THE UNIVERSITY
of EDINBURGH

THE UNIVERSITY
of EDINBURGH

Targeted genome editing to replace nuclear-encoded Rubisco in higher plants

Sophie Donovan

Doctor of Philosophy

University of Edinburgh

2019

Declaration

I declare that this thesis was composed by myself and that the work presented here is my own unless indicated otherwise. This work has not been submitted for a degree or any other professional qualification.

Sophie Donovan
16th March 2020

Acknowledgements

During my four years at the University of Edinburgh I have been fortunate enough to be surrounded by people that have supported me through a challenging time. There are no words to express my gratitude and appreciation for the people that have helped me reach this point, but I will give it a go.

First, I want to thank my supervisor Alistair, for supporting and sometimes objecting to my craziest ideas. Our discussions helped me to become a better scientist, encouraged me to keep going and pushed me to write more. I would also like to thank my thesis committee members, Karen and Steven, for valuable input, advice and encouragement. Thank you to all of the staff who keep things moving in Rutherford, particularly to Billy and Pat for keeping my tobacco plants happy and healthy.

The close friendships that I have formed in the McCormick lab have made my time here truly enjoyable. Thank you, Nicky, for helping me to avoid protein related meltdowns and for awesome life advice; Livia for being so wise, supporting me and looking after the afternoon chocolate supply; Chris for always criticising my work; Ale por las risas, las lecciones de español y todas las cervezas.

Finally, thanks PK, you are the best colleague, collaborator, flatmate and friend who has stuck with me through the best and worst times. Thank you for keeping me company during eighteen-hour time courses and the evenings drinking wine and talking about Rubisco and other important things in life. It is not possible to express how much I appreciate your support, but I am truly thankful for our friendship.

I am extremely grateful to my friends and family for their continuous encouragement. Thank you, Susie, Olivia, Emily, Sally, Eric, Gaby, Capilla, Samuel, Ben, Greg, Stef, and Sean for the breaks, dancing and company that I needed to not feel so lonely. Thanks to my grandparents who have always been there to listen. Finally, thank you to my parents and sister, Lesley, Chris, and Lily, who have always encouraged me and advised me to do what I enjoy.

Lay Summary

Plants, algae and some bacteria use sunlight, water, and the gas carbon dioxide (CO_2) to produce energy. This process is called photosynthesis and it is essential to produce the sugars that are required for growth. Photosynthesis is surprisingly inefficient in most crop plants. It is important to study photosynthesis (1) to understand why photosynthesis can be inefficient and (2) to identify ways to improve photosynthesis and increase the yield of crops.

One of the first stages of photosynthesis in plants is the capture of CO_2 from the atmosphere. A protein called Rubisco captures CO_2 and converts it into a molecule that is later used to synthesise sugar. However, Rubisco cannot tell the difference between CO_2 and O_2. Sometimes Rubisco captures O_2 instead of CO_2 and produces a different molecule, which reduces the efficiency of photosynthesis.

We are interested in identifying why Rubisco proteins from some species are better at discriminating between CO_2 and O_2 than others. To explore this, it is important to understand the composition of the Rubisco protein. First, this work investigates the composition of Rubisco in a model crop species (tobacco). Second, plants with lower amounts of native Rubisco were made. Reducing the amount of native Rubisco is useful for future studies to replace it with Rubisco from another species. This will allow us to explore the impact of Rubisco with different CO_2 and O_2 specificities on plant growth and photosynthesis.

Abstract

Improving the efficiency of the Rubisco enzyme (ribulose-1,5-bisphosphate carboxylase/oxygenase) is a key strategy to enhance photosynthesis and yields in crops. Rubisco catalyses net CO_2 assimilation in all photosynthetic organisms but is slow and cannot fully discriminate between O_2 and CO_2. The small subunit of Rubisco (SSU) can play an important role in determining catalytic rates. However, SSUs are encoded by large, nuclear *rbcS* gene families. *Nicotiana tabacum* (tobacco) is a model often used for testing Rubisco engineering approaches. Nevertheless, the *rbcS* family in tetraploid tobacco remains poorly understood. First, this work characterised the *rbcS* gene family to identify the major isoforms that contribute to the tobacco Rubisco enzyme. The regulation of individual *rbcS* isoforms in response to light quantity and quality was also explored. Second, a strategy to knock-out the major *rbcS* isoforms using an RNA-guided endonuclease (RGEN) was established. A tobacco mutant with *ca.* 5% of wild-type Rubisco content was successfully generated as a tool for future studies to engineer Rubisco. Finally, an approach to examine the impact of heterologous SSUs on growth and photosynthesis in Arabidopsis and tobacco was also explored. These findings contribute to efforts to engineer Rubisco by providing a platform to express non-native SSUs in a model crop species.

List of Abbreviations

2-PG	2-phosphoglycolate
3-PGA	3-phosphoglycerate
A	photosynthesis
A_{max}	maximum rate of photosynthesis
ANOVA	analysis of variance
BL	blue light
BSD2	bundle sheath defective 2
C_a	external CO_2 concentration
CABP	carboxyarabinitol bisphosphate
CCM	carbon-concentrating mechanism
C_i	internal CO_2 concentration
CRISPR	clustered regularly interspaced short palindromic repeats
CKABP	3-keto-2'-arbinitol-1,5-bisphosphate
DAG	days after germination
DSB	double-stranded break
EMS	ethyl methanesulfonate
F_{LUC}	firefly luciferase
FRL	far-red light
G3P	glycerate-3-phosphate
gRNA	guide RNA
HDR	homology-directed repair
HSD	honestly significant difference
HSP	heat shock protein 18.2
IBA	indole-3-butyric acid
Indel	insertion or deletion
J_{max}	maximum electron transport rate
k_{cat}^c	turnover rate for CO_2
LB	luria broth
LSU	large subunit
MOPS	3-(N-morpholino)propanesulfonic acid
NGS	next generation sequencing
NHEJ	non-homologous end joining

$\mathbf{N_{LUC}}$	nano luciferase
PAM	protospacer adjacent motif
PAR	photosynthetically active radiation
PCR	polymerase chain reaction
$\boldsymbol{Q}$	irradiance
qPCR	quantitative polymerase chain reaction
Rubisco	ribulose-1,5-bisphosphate carboxylase/oxygenase
RuBP	ribulose-1,5-bisphophate
RAF1	Rubisco accumulation factor 1
RAF2	Rubisco accumulation factor 2
RGEN	RNA-guided endonuclease
RL	red light
RL-FRL	red light and far-red light
$\boldsymbol{S_{c/o}}$	Rubisco specificity factor
SDS-PAGE	sodium dodecyl sulphate-polyacrylamide gel electrophoresis
SSU	small subunit
$\boldsymbol{T_{leaf}}$	leaf temperature
TP	transit peptide
Ubi10	ubiquitin 10
$\boldsymbol{V_{c,max}}$	maximum carboxylation rate of Rubisco
Γ	CO_2 compensation point
μ	micro
$\boldsymbol{\varepsilon_c}$	maximum energy conversion efficiency of photosynthesis

Table of Contents

Chapter 1

Literature Review

1.1 Rubisco catalysis

Ribulose-1,5-bisphosphate carboxylase/oxygenase (Rubisco) is the primary enzyme responsible for biological CO_2 assimilation in photosynthetic organisms. Rubisco is described as the most abundant protein on Earth and accounts for at least half of the total protein in leaves. More than 90% of Rubisco on Earth is found in the leaves of plants (Bar-On and Milo, 2019). Despite a central role in CO_2 fixation, Rubisco is considered a major rate-limiting step in photosynthesis because it is a bifunctional enzyme with a competitive oxygenase activity and slow carboxylase activity. The oxygenase reaction reduces photosynthetic efficiency in C3 crops, including rice and wheat (Ogren, 1984; Sharkey, 1988). In C3 plants, replacing Rubisco with faster or more specific enzymes is an important strategy to increase photosynthesis in crops (Parry et al., 2013).

Rubisco catalyses the carboxylation of ribulose-1,5-bisphosphate (RuBP) to eventually yield two molecules of 3-phosphoglycerate (3-PGA) (Fig.1–1) (Taylor and Andersson 1997) initiating a series of steps resulting in sugar synthesis and the regeneration of RuBP. The reaction occurs as follows: (1) enolisation of RuBP by deprotonation at C3 forming a 2,3-enediol intermediate, (2) carboxylation of 2,3-enediol forming 3-keto-2'-arbinitol-1,5-bisphospate (CKABP) and hydration to form a gem diol intermediate, (3) cleavage of the C2-C3 bond forming one molecule of glycerate-3-phosphate (G3P) in a carbanion form and one molecule of 3-PGA, and (4) protonation of the carbanion to yield another molecule of 3-PGA.

Oxygenation of RuBP forms one molecule of 3-PGA and one molecule of 2-phosphoglycolate (2-PG). 2-PG is re-salvaged to 3-PGA via the photorespiratory salvage pathway

(photorespiration), which results in the loss of previously fixed CO_2. After photorespiration, the maximum energy conversion efficiency (ε_c) of photosynthesis is halved in C3 plants (Zhu, et al., 2008; 2010). CO_2 and O_2 compete for the active site of Rubisco and oxygenation decreases photosynthetic efficiency by *ca.* 25% (Sharkey, 1988).

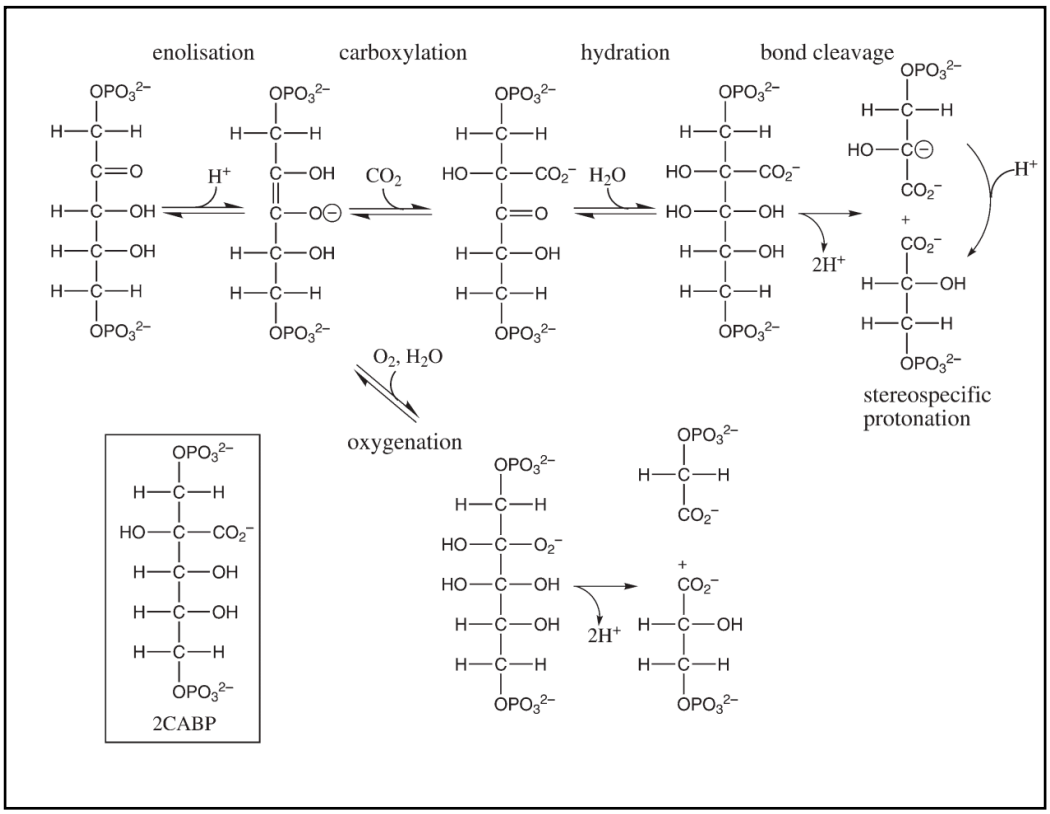


Figure 1–1. Rubisco catalysed carboxylation and oxygenation of RuBP. Figure from (Andersson, 2008).

The dual catalytic activities of Rubisco are defined by the turnover rate for O_2 (V_o or k_{cat}^o) and CO_2 (V_c or k_{cat}^c) and the Michaelis Menten constants for CO_2 and O_2 (K_c and K_o, respectively). The CO_2/O_2 specificity factor ($S_{C/O} = V_cK_o/V_oK_c$) describes the relationship between the carboxylation and oxygenation (Laing *et al.,* 1974; Jordan & Ogren, 1984). Cyanobacteria, algae, and some plant species (C4 and CAM) have evolved carbon-concentrating mechanisms (CCMs) that limit oxygenation by increasing the availability of CO_2. Rubisco enzymes from species with CCMs are usually faster and have lower specificities for CO_2.

Average $S_{c/o}$ values (at 25°C) are highest in marine red algae (185) followed by terrestrial C3 plants (85), while C4 and green algal enzymes have lower specificities (79 and 59 respectively). Until recently, an inverse correlation between $S_{c/o}$ and k_{cat}^{c} had been observed for Form I Rubisco enzymes from plants and algae (Tcherkez *et al.,* 2006). The evolution of Rubisco could be limited by physio-chemical constraints (Tcherkez *et al.,* 2006; Savir *et al* 2009). However, this correlation did not reflect the diversity of Rubisco enzymes. Eleven enzymes from diatom species lack the trade-off between carboxylation rates and specificity (Young *et al.,* 2016). Therefore, it could be possible to engineer a more specific Rubisco enzyme while maintaining high carboxylation rates (Hanson, 2016).

1.2 Opportunities for improvement

Expressing faster or more specific Rubisco enzymes in the leaves of C3 crops has the potential to increase photosynthesis (Parry et al., 2013). For example, Rubisco isolated from a non-green alga (*Griffithsia monilis*) has a higher specificity than the tobacco (*Nicotiana tabacum*) form but a similar k_{cat}^{c} (Whitney *et al.,* 2001). Replacing the C3 enzyme with a more specific form could increase photosynthesis by *ca.* 27% in tobacco leaves (Zhu *et al.,* 2004). Rubisco enzymes from different plant species are also a source of kinetic diversity. High temperatures reduce the specificity of Rubisco and can increase photorespiration (Jordan & Ogren, 1984; Brooks & Farquhar, 1985; Sage, 2002). As a consequence, some C3 species from arid environments have higher $S_{c/o}$ values (Galmes *et al.,* 2005). Expressing a Rubisco enzyme that has a high specificity (e.g., from *Limonium gibertii* (110.5)) could increase photosynthesis in tobacco by 5-30% depending on the compromise to k_{cat}^{c} (Galmes *et al.,* 2005). Further gains could be achieved by expressing different forms of Rubisco in different environments or developmental stages (Zhu & Long, 2009). For example, in light saturated leaves photosynthesis is Rubisco-limited; therefore, the benefits of an enhanced k_{cat}^{c} might compensate for lower $S_{c/o}$. However, in shaded leaves, RuBP regeneration is limiting so a higher $S_{c/o}$ would maximise photosynthesis by reducing photorespiration. Expressing a faster

Rubisco at the top of a C3 crop canopy could reduce Rubisco-limited photosynthesis in sunlit leaves. In shaded leaves at the bottom of the canopy, expressing a more specific Rubisco could increase net carbon gain by reducing photorespiratory loss. Manipulating tobacco to express different forms of Rubisco in specific leaves in the canopy could improve photosynthesis by 31% (Zhu & Long, 2009).

1.3 Rubisco structure and assembly

Rubisco (Form I Rubisco) in higher plants and algae is composed of eight large subunits (LSUs) (*ca.* 55 kDa each) and eight small subunits (SSUs) (*ca.* 15 kDa each) (Fig. 1–2). Form II enzymes, which are found in some prokaryotes, function solely as LSU dimers. Furthermore, some archaea have Form III enzymes that exist as LSU dimers in a decameric arrangement (Form II and III structures are reviewed in detail in (Tabita et al. 2008)). The LSU consists of a small N-terminal domain and a larger C-terminal domain that forms an α/β-barrel (Whitney et al., 2011). Two active sites are formed at the interface between LSU dimers with most of the active site residues contributed by loops connecting the α/β-barrel (Andersson et al., 1989). Two groups of four SSUs cap the top and bottom of eight LSUs forming an L_8S_8 complex (*ca.* 560 kDa).

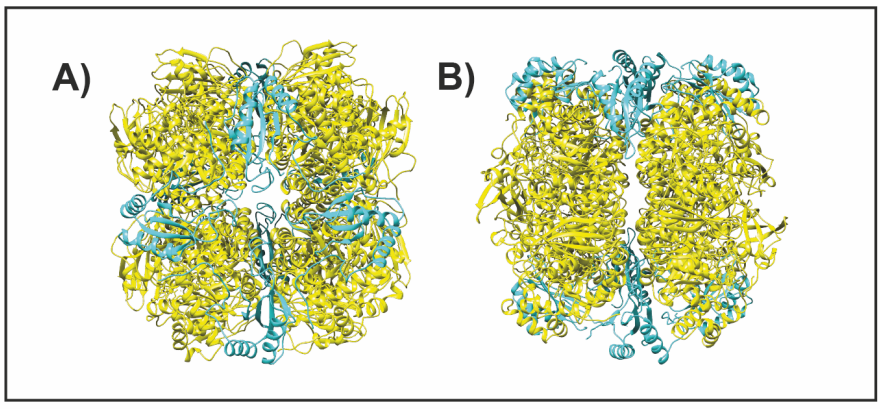


Figure 1–2. The structure of Form I Rubisco (A) Four small subunits (SSUs) (blue) cap the top and bottom of the large subunit (LSU) core (yellow). **(B)** L_8S_8 complex viewed from the side. Structure shows spinach Rubisco (Protein Databank: 1RCO).

Mutagenesis of eukaryotic Form I Rubisco in plants is challenging due to the complex requirements for assembly and biogenesis. The LSU gene (*rbcL)* is chloroplast encoded and the SSU genes (*rbcS)* are nuclear-encoded by a family of isoforms. Currently, replacement of native LSUs with plant-like sequences is limited to *Chlamydomonas reinhardtii* and tobacco (Kanevski et al. 1999; Sharwood et al. 2008; Spreitzer and Salvucci 2002; Whitney and Sharwood 2008). Furthermore, folding of LSUs and assembly of the L_8S_8 complex is mediated by multiple chaperones proteins that have co-evolved with Rubisco in a species-specific manner (Hauser et al., 2015). Although it is possible to express prokaryotic Rubisco enzymes in *E. coli* (Gatenby et al., 1985; Gutteridge et al. 1984), the requirement for additional assembly factors long prevented the assembly of plant Rubisco (Cloney *et al.*, 1993).

Establishing the assembly requirements of eukaryotic Rubisco is important for heterologous expression studies. Recently a functional Arabidopsis Rubisco enzyme was assembled in *E. coli* by co-expressing LSUs and SSUs with five cognate chaperones (Aigner et al., 2017). In plants, folding of the large subunit is mediated by a chaperonin (Cpn60) and associated co-factors (Cpn10/Cpn20) (Bracher et al., 2011; Tsai et al., 2012) (Fig. 1–3). The assembly of LSU dimers is mediated by the co-ordinated or parallel actions of RbcX and a species-specific rubisco accumulation factor 1 (RAF1) (Aigner et al., 2017; Hauser et al., 2015; Whitney et al., 2015). Eight bundle-sheath-defective-2 (BSD2) proteins interacts with a single LSU_8 core likely forming an end-state assembly intermediate before displacement of BSD2 with eight SSUs (Aigner et al. 2017). SSUs are synthesised as a precursor in the cytosol and translocated to the chloroplast where transit peptide is cleaved to form a mature SSU. An additional protein, RAF2, is essential for enzyme assembly; however, the role of RAF2 has not been clarified.

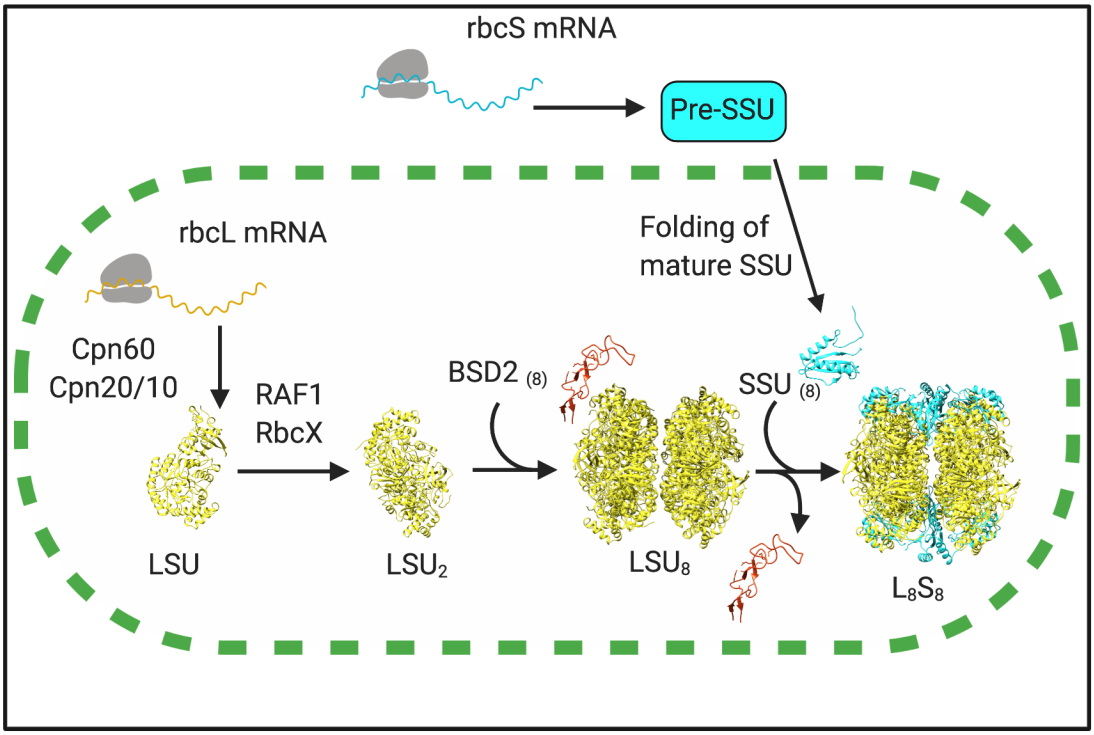


Figure 1–3. Chaperones involved in the assembly of Form I Rubisco in plants. The precursor for the small subunit (SSU) is translocated to chloroplast and the signal peptide is cleaved and folded in to a mature SSU. The large subunit (LSU) is folded by the chaperonin 60 (Cpn60) complex and co-factors (Cpn20/Cpn10). Dimers of LSU assemble via rubisco accumulation factor 1 (RAF1) and RbcX. Eight bundle sheath defective 2 (BSD2) proteins (PD:6EKB) stabilise the LSU core before displacement with mature SSUs to form the complete L_8S_8 complex (PD: 1RCO).

1.4 Genetic manipulation of Rubisco

As previously mentioned, three predominant genetic manipulation techniques have been used in combination with phylogenetic studies to identify residues that influence the catalysis and assembly of Rubisco: (1) directed mutagenesis of *Synechococcus* and *Rhodospirillum rubrum* enzymes expressed in *E. coli*; (2) screening and recovery of *C. reinhardtii* Rubisco mutants that can survive on an alternative carbon source in the absence of photosynthesis; and (3) plastid transformation of tobacco with mutated and heterologous LSU sequences.

1.4.1 Large subunit studies

The first two approaches confirmed that the residues in the active site are essential for maximal carboxylase activity (Fitchen et al., 1990; Harpel and Hartman, 1994). Furthermore, residues that interact with a transition state analog that binds the active site (2-carboxyarabinitol-1,5-bisphosphate (CABP)) are highly conserved and cannot account for the variation in catalytic parameters (Spreitzer and Salvucci 2002). Residues surrounding and distant from the active site that influence enzyme stability and $S_{C/O}$ have been identified (reviewed in Spreitzer and Salvucci, 2002; Andersson and Backlund, 2008). These include structural elements close to active site residues, the N-terminal domain, and the interface between LSUs and SSUs (Spreitzer and Salvucci, 2002; Kapralov and Filatov, 2007). The final approach is essential to establish compatibility between plant assembly chaperones and heterologous Rubisco enzymes, in addition to understanding the effect of altered catalytic properties on plant physiology (Andrews and Whitney, 2003).

Precise replacement of *rbcL* by homologous recombination in plants is well established in tobacco (Svab and Maliga 1993). Tobacco Rubisco can be replaced with more simple Rubisco enzymes that function as LSU dimers from *Methanococcoides burtonii* and *Rhodospirillum rubrum* (Alonso et al., 2009; Whitney and Andrews, 2003; Wilson et al., 2016). However, Form I enzymes from more distant species, including red-algae and cyanobacteria, fail to assemble in tobacco chloroplasts without the co-expression of a cognate assembly factor (Lin et al., 2014; Lin and Hanson, 2018). The requirement for species-specific assembly factors can also limit the biogenesis of chimeric enzymes formed of tobacco SSUs and heterologous LSUs from other plant species. Functional Rubisco enzymes can be produced from tobacco SSUs and LSUs from plant species including *Solanum lycopersicum* (tomato), *Healianthus annuus* (sunflower) and *Flaveria* (Kanevski et al., 1999; Sharwood et al., 2008; Whitney et al., 2011; Zhang et al., 2011). However, the capacity to produce wild-type amounts of chimeric Rubisco enzymes is reduced despite sharing the same regulatory sequences and high sequence

similarity between LSUs. Expression of RAF1 improves the biogenesis of chimeric Rubisco formed of Arabidopsis LSUs and tobacco SSUs. (Whitney et al., 2015). Therefore, continuing to identify RAF1 orthologs and other chaperone proteins can improve efforts to assemble chimeric enzymes in plants.

1.4.2 Small subunit studies

SSUs are more diverse than LSUs, and although they are known to affect catalysis their role remains somewhat enigmatic (Spreitzer, 1999; Spreitzer, 2003). Chimeric Rubiscos formed of native LSU and plant SSUs assemble in *C. reinhardtii* mutants lacking endogenous *rbcS* (Khrebtukova and Spreitzer, 1996; Genkov et al., 2010). Introducing a spinach (*Spinacia oleracea*), Arabidopsis or sunflower SSU increased $S_{c/o}$ by 3-11%. However, expression of the plant SSUs resulted in loss of the pyrenoid, a key element of the algal CCM, so improvements in Rubisco catalysis did not translate to increased photosynthetic rates.

Major structural variation exists in the SSU βA-βB loop that forms an interface with LSUs (Esquivel *et al.,* 2013). On average, prokaryotes and non-green algae have 10 residues in the loop, while plants and green algae have 22 and 28 residues respectively (Spreitzer & Salvucci, 2002). Variation in the SSU βA-βB loop may account for differences in the kinetic characteristics of divergent enzymes (Karkehabadi *et al.,* 2005). Substitution of the algal βA-βB loop with the shorter loops from spinach produced a functional chimeric enzyme with altered catalytic properties. Furthermore, modifying the algal SSU with the βA-βB loops from *S. elongatus* loop decreased $S_{c/o}$ and k_{cat}^{c} towards values for cyanobacterial enzymes. In contrast, the longer spinach loop resulted in a decrease in k_{cat}^{c} but did not affect $S_{c/o.}$ Co-engineering of the algal βA-βB loop and LSU residues to those characteristic of C_3 Rubisco enzymes increases $S_{c/o}$ and decreased k_{cat}^{c}, mimicking the catalytic properties of C_3 Rubisco (Spreitzer e*t al.,* 2005).

Expressing chimeric enzymes with non-native SSUs in plants is challenging owing to the presence of native SSUs. Furthermore, it can be difficult to determine the contribution of non-native SSUs owing to the methods used to measure Rubisco content. Rubisco content can be determined by SDS-PAGE and immunoblotting or binding of ^{14}C-CABP to the active sites of the enzyme (Kubien *et al*., 2011). Immunoblotting depends on the use of a quantified protein standard but can result in the overestimation of Rubisco content owing to the large amount of Rubisco in leaves. Therefore, the ^{14}C-CABP assay is the preferred method for detailed Rubisco studies (Kubien *et al*., 2011). Previous work mostly resulted in chimeric forms of the enzyme that were formed from both native and heterologous SSUs. For example, expression of a *rbcS* gene from pea in Arabidopsis resulted in heterologous pea SSUs contributing to 15-18% (*ca*. 1 pea SSU per holoenzyme) of the total SSU content as determined by SDS-PAGE (Getzoff *et al*., 1998). However, the carboxylase activity of the chimeric enzyme was reduced relative to wild-type Rubisco, possibly due to an incompatible interaction between the Arabidopsis and pea SSUs.

Supressing the expression of native SSUs is important to examine the effect of heterologous SSUs on growth and photosynthesis. A heterologous SSU from *Sorghum bicolor* was overexpressed in the leaves of rice plants and accounted for 79% of total *rbcS* in one transgenic line (Ishikawa *et al.,* 2011). Incorporation of the sorghum SSU significantly increased k_{cat}^{c} and decreased $S_{c/o}$, shifting the kinetic characteristics of Rubisco towards that of a C4-like enzyme. Mutant lines with reduced amounts of native SSUs offer another approach for heterologous SSU expression. An Arabidopsis T-DNA insertion line (*1a3b*) had *ca.* 30% of native *rbcS* expression (Izumi et al., 2012). Complementation of *1a3b* with a heterologous SSU from *C. reinhardtii* restored Rubisco content to 65% of wild-type levels (Atkinson et al., 2017).

Introducing divergent SSUs or LSUs can have a significant effect on plant growth and photosynthesis. Nevertheless, these studies are limited by lack of a suitable platform for co-expression of engineered SSU and LSUs in plants.

1.5 The *rbcS* gene family

In plants, SSUs are encoded by families of *rbcS* isoforms that usually produce identical peptides. However, in cyanobacteria, a single *rbcS* gene is co-transcribed with *rbcL* from the same operon. Therefore, the migration of *rbcS* genes to the nucleus occurred before the divergence of land plants. In species with nuclear-expressed *rbcS*, the size of the *rbcS* family ranges from 2 (*C. reinhardtti*) to 22 (*Triticum aestivum*) members (Spreitzer, 2003). The evolution of *rbcS* families in plants suggests that natural selection has favoured the production of more Rubisco to compensate for the catalytic inefficiencies of Rubisco in plants (Spreitzer, 2003). Little is known of the functional significance of individual *rbcS* isoforms within a particular family but often one or two isoforms account for the majority of *rbcS* transcripts. Furthermore, individual isoforms are expressed differentially during development, in different organs, and in response to high temperatures and light (Wanner & Gruissem, 1991, Dedonder *et al.,* 1993, Sawbridge et al., 1996, Yoon *et al.,* 2001, Suzuki *et al.,* 2009). It is likely that differential expression of *rbcS* isoforms operates to adapt Rubisco content, and possibly catalytic properties, to the environment (Yoon et al., 2001).

Phylogenetically distinct *rbcS* isoforms that produce faster and less specific Rubisco enzymes have been identified in land plants (Morita et al., 2016; Laterre et al., 2017; Pottier et al., 2018). The isoforms (referred to as *rbcS-T*) are expressed in non-photosynthetic tissues (e.g. trichomes and roots) and are genetically distinct from mesophyll-expressed isoforms (*rbcS-M*) within the same species. It is possible that *rbcS-T* isoforms are involved in specialised metabolic functions owing to exclusive expression in non-photosynthetic tissues. Overexpression of a *rbcS-T*-like isoform (*OsrbcS1*) in the leaves of rice altered the catalytic

properties of Rubisco to C4-like (Morita et al., 2014). Similarly, a faster Rubisco is produced from SSUs derived from a *rbcS-T* isoform that is expressed exclusively in tobacco trichomes than tobacco *rbcS-M* (Laterre et al., 2017). Replacement of *rbcS-M* with *rbcS-T* isoforms could produce a faster Rubisco without the need to express cognate chaperone proteins. However, it would still be necessary to supress the expression of *rbcS-M* isoforms.

As previously mentioned, it is difficult to engineer a knock-out of *rbcS* in plants. This is largely because of the size of the family, which often includes linked isoforms that cannot be simultaneously knocked out by T-DNA insertion and crossing. However, RNA-guided endonucleases (RGENs) including Cas9 (CRISPR-associated protein 9) (also known as CRISPR/Cas9) can be used to edit multiple and closely linked genes simultaneously, including the *rbcS* family.

1.6 Editing plant genomes with RNA-guided endonucleases

Over the last six years RGENs have enabled the precise editing of genomes in both crop and model species (Schiml & Puchta, 2016; Raitskin & Patron, 2016). The type II prokaryotic CRISPR (clustered regularly interspaced short palindromic repeats)/Cas adaptive immune system was repurposed to facilitate targeted RNA-guided DNA cleavage (Cong et al., 2013). The *Cas9* endonuclease from *Streptococcus pyogenes* was one of the first RGENs to be domesticated and used for mutagenesis in plants (Jinek et al., 2012; Li et al., 2013; Nekrasov et al., 2013). Although this section will focus on *SpCas9*-mediated approaches it is important to note that the subsequent development of other Type II RGENs have expanded the capacity to manipulate plant genomes (reviewed in Shmakov et al., 2017; Khumsupan et al., 2019).

SpCas9 is guided to a DNA-target by a synthetic guide RNA (gRNA) composed of a short (*ca.* 20 nucleotides) sequence that is complementary to the target site and next to a protospacer adjacent motif (PAM) (5'-NGG-3') (Jiang and Doudna, 2017). Complementary binding triggers a conformational change in the two nuclease domains of *SpCas9* (RuvC and HNH)

that induce a double-stranded break (DSB) 3-4 bp proximal to the PAM. Mutations are introduce by repair of the DSB. In plants, the predominant repair pathway is non-homologous end joining (NHEJ). NHEJ is error-prone and often results in the insertion or loss of nucleotides at the DSB site. More precise repair homology-directed repair (HDR) can occur when an exogenous repair template is provided (Shi *et al.,*2016). However, the efficiency of HDR varies between different plant species (Khumsupan et al., 2019).

Although the generation of NHEJ-based mutations is efficient, characterising the heritability of mutations remains a considerable challenge for RGEN approaches. Homozygous or bi-allelic mutations have been reported in a single generation in *Oryza sativa* (rice) and tomato (Zhang *et al.,* 2014; Brooks *et al.,* 2014). However, mutations occuring in somatic cells (chimeric) are more frequent. It is difficult to confirm if mutations are heterozygous and bi-allelic or chimeric without examining the segregation of mutations in later generations (Feng *et al.,* 2013; Jiang *et al.,* 2014). Therefore, it is desirable to screen for mutated lines that have the *SpCas9* transgene segregated out. Furthermore, transgene-free edited plants that could be produced by breeding are not subject to genetically modified organism (GMO) regulations in countries including the United States (USDA).

A key advantage of RGENs is the capacity to target multiple genes simultaneously in a single transformation event. Therefore, RGENs are well suited for editing polyploid genomes or gene families. Orthologous genes with a high nucleotide identity can be edited using a single gRNA sequence that is complementary to multiple targets (Endo et al., 2015; Sun et al., 2018). Alternatively, genes with lower nucleotide identity can be targeted by expressing multiple gRNA in a single vector (Lowder et al., 2015; Tang et al., 2019). The mutation of four genes involved in carotenoid biosynthesis has been reported in the T_0 generation in tomato (Li et al., 2018). Furthermore, homozygous sextuple mutants have been obtained in the T_3 generation of

Arabidopsis (Zhang et al., 2016). These studies highlight the potential for RGEN editing of multiple *rbcS* isoforms in plants.

1.7 Thesis Aims

An understanding of effect of Rubisco enzymes with altered catalytic properties on plant growth and photosynthesis requires a platform for simultaneous expression of heterologous LSUs and SSUs. Despite well-established protocols to replace the tobacco *rbcL* gene, the presence of the native *rbcS* family prevents the efficient expression of a heterologous Form I enzyme in tobacco. Furthermore, there is little understanding of the individual *rbcS* isoforms that contribute to tobacco Rubisco. First, this work aims to characterise the tobacco *rbcS* family by identifying the genomic location and differential expression of individual isoforms. Based on the *rbcS* families that have been characterised to date, two or three *rbcS* isoforms account for the majority of *rbcS* expression in plants. Therefore, we expect that only a few isoforms will account for the majority of *rbcS* expression in tobacco. Second, this work aims to develop a strategy to remove and replace the native SSUs in tobacco. With recent advancements in genome editing, the removal and replacement of multiple *rbcS* isoforms is now a feasible goal. Antisense studies have shown that decreasing the content of Rubisco to less than 40% of wild-type severely impairs photosynthesis and growth of tobacco plants (Masle et al., 1993; Quick et al., 1991a; Quick et al., 1991b; Stitt et al., 1991). Similarly, we expect that knocking out the most highly expressed *rbcS* isoforms in tobacco will produce a similar phenotype that can be used to test the effect of non-native SSUs on plant photosynthesis and growth.

Chapter 2

Material and Methods

2.1 Identification of tobacco *rbcS* isoforms

2.1.1 Database searching

Partial genomic DNA coding sequences for the thirteen known tobacco *rbcS* isoforms (Gong et al., 2014) (Table 1) were used to BLAST search three tobacco genome assemblies on the Sol Genomics database (https://solgenomics.net/) (Sierro et al., 2014). Genomic regions with >99% identity to at least 75% of a single input query were selected as putative *rbcS* isoforms. Putative full-length isoforms were aligned to the previously characterised full-length coding sequence from tobacco cv. Samsun (X02353.1) using EMBOSS Needle (EMBL-EBI) to identify the predicted transcriptional start and termination sites. Partial coding sequences mapped to chromosomes by BLAST searching a second genome assembly on the Sol Genomics website (Edwards et al., 2017).

2.1.2 Validation of full-length coding sequences

The predicted full-length coding sequences were used as input queries for Primer-BLAST (NCBI). Primer pairs specific to the target queries were chosen for isoform-specific PCR (Table 2-1). Full-length isoforms were amplified using Q5 high-fidelity polymerase (NEB, UK) and the PCR products sequenced using both forward and reverse primers. The sequencing reads were aligned with the partial-coding sequences and alignments with ≥99% identity were considered correct.

2.1.3 Bioinformatic methods

Sequences derived from the methods described in sections 2.1.1 and 2.1.2 were used for multiple sequence alignments with Clustal Omega (EMBL-EBI) using default parameters. Multiple sequence alignments were used to produce phylogenetic trees with Simple Phylogeny (EMBL-EBI) using the neighbour-joining method that were edited using iTOL (https://itol.embl.de).

2.2 DNA methods

2.2.1 DNA extraction

Genomic DNA was extracted from *ca.* 10 mg of leaf tissue using cetyl trimethylammonium bromide (CTAB) extraction. Frozen tissue was lysed in a TissueLyser using a single 3mm tungsten bead and 500 µL of extraction buffer (0.2M Tris-HCl, 0.05M EDTA, 2M NaCl, 2% CTAB, pH 7.5) added. Lysed samples were heated to 65°C for 1 hour and 250 µL of chloroform added before vigorous shaking for 5 minutes. Samples were spun at 14000 rpm for 10 minutes and 250 µL of the supernatant was added to an equal volume of propanol and incubated at -20°C for two hours. DNA was pelleted by centrifugation at 14000 rpm for 30 minutes and washed in 800 µL of ethanol. Pellets were left to dry for at least two hours before being dissolved in 50 µL of sterile distilled water.

2.2.2 Polymerase Chain Reaction

Q5 High-Fidelity Master Mix (#M0515, NEB) was used for amplifying tobacco *rbcS* sequences and cloning in accordance with the manufacturer's instructions. Forward and reverse primers were used at a final concentration of 0.5 µM each and the following thermal cycling parameters were used unless otherwise specified (98°C 30 sec; 30 x 98°C 5 sec, 58–66°C 10 sec, 72°C 20–30 sec/kb; 72°C 2 min). For colony PCR, MyTaq Red Mix (#25043,

Bioline) was used with 0.4 μM of each primer and the following cycling conditions unless otherwise specified (95°C 1 min; 25 x 95°C 15 sec, 58–66°C 15 sec, 72°C 10 sec/kb).

2.2.3 Sequencing

Sanger sequencing was performed using the Big Dye Terminator v3.1 Cycle Sequencing Kit (#4337455, ThermoFisher Scientific) and analysed by Edinburgh Genomics (Edinburgh, UK).

2.2.4 Mutation screening

Mutations in *rbcS-T1* were detected by gene-specific PCR and sequencing (Table 2-1 F1 and R1) using a genomic DNA template. Mutations in *rbcS-S1* were detected by gene-specific PCR and sequencing using a cDNA template. Mutations were identified by pairwise sequence alignment (EMBL-EBI NEEDLE) with the wild-type genomic DNA sequence (Madeira et al., 2019). Mutation frequencies were determined from the sequencing chromatograms using the tool TIDE (https://tide.deskgen.com) (Brinkman et al., 2014).

2.2.5 Table of primers for amplifying *rbcS* genes

Table 2-1. Primers for amplifying and sequencing tobacco rbcS genes. ID refers to GenBank accession number and T_a shows the actual annealing temperature used for amplification. Table continued on next page.

Gene	ID	No.	Primer	T_a (°C)
rbcS1a/b	KM025316.1	F1	TAGGGTGGTGGGCAACTATG	64
	KM025317.1	R1	TTCAAACAAACTGCCCCTAAA	64
		F2	GCAGCAGTTGCCACCCGCAG	60
		R2	GCAATGAAACTGATGCACTGCACTT	60
rbcSS2	KM025319.1	F1	GCCCAAGGAGATTCAAACAA	62
		R1	TGGGGAAGCTATGTGAAACC	62
		F2	GCAGCAGTTGCGACCGGCGCT	58
		R2	GCGATGAAACTGATGCAT	58
rbcSS3	KM025321.1	F1	TGCAAACAAACTTTCCCTGA	62
		R2	GCTTGATTTGTGTCCGTTGA	62
		F2	GCAGCAGTTGCGACCGGCA	60
		R2	GCGATGAAACTAATGCAC	60
rbcSS4	KM025323.1	F1	CCCATTCACAAATTATGTCAGG	62
		R1	GAAAAACCAAAACAGTTTCTCCA	62
		F1	GCAGCAGTTGCCACTGGCGCT	60
		R2	Identical to S2_R2	-
rbcSS5	KM025325.1	F1	CAAGACTCCGGGACAGAAAG	62
		R1	AACGGCTACCATTCCTCTCA	62
		F2	GCAGCAGTTGCCACCGGCGCC	60
		R2	GCGATGAAACTGATGCAC	60

Table 2-1 (continued). Primers for amplifying and sequencing tobacco rbcS genes. ID refers to GenBank accession number and T_a shows the actual annealing temperature used for amplification.

Gene	ID	No.	Primer	T_a (°C)
rbcST1	KM025327.1	F1	TAGGGTGGTGGGCAACTATG	64
		R2	CGGCACAAGAATGTGAAACA	64
		F2	GCAGCAGTTGCCACTCGCAC	60
		R2	Identical to S1_R2	-
rbcST2	KM025329.1	F1	TCAAATAACCCTCTTGAAAGCAA	60
		R1	GTGTAATGTCAGGGGCCAAA	60
		F2	GCAGCAGTTGCGACCGGCGCC	60
		R2	GCAATAAAACTGATGCAC	60
rbcST4	KM025334.1	F1	GGAAAACCAAAACAGTTTCTCC	62
	KM025335.1	R1	GGTCGCTGCTAAAATAGTCACA	62
		F2	GCAGCAGTTGCCACCGGCGCT	60
		R2	Identical to S2_R2	-
rbcST5	KM025337.1	F1	CCGTATGTGCAACTTCATCG	57
		R1	ACTTCCCCCGAAGACATAGG	69
		F2	Identical to T2_F2	-
		R2	GCAATAAAACTGATGCAC	60

2.2.6 Table of primers for screening transgenic plants

Table 2-2. Primers for screening transgenic plants. Primers used to confirm transgene integration with T_a shows the actual annealing temperature used for amplification.

Gene	Sequence	T_a (°C)
	CGCTAATCTTGCAGGTAGCC	58

spCas9	AGCCCCGTAATTGACTGATG	
CrrbcS2	GCCGAGAGCGATAAAGCCTA CGGCCTCTGTACAAGGAACC	60
HarbcS	CCACCACTTGGACTCAAGAAG CTGGTCGGGATGCTATGAAC	60
LgrbcS	ACCCGAGGGTTTGAAAAAGT TACACTGGACCTGACGCTTG	60
SbrbcS	AAGTTTGGCCAGCTTATGGA CAAGAATTCTAACATAAGCATCTGGA	60

2.3 RNA methods

2.3.1 RNA extraction

Total RNA was isolated using the RNeasy Mini Kit (#74104, QIAGEN) and RNA was treated with RNase free DNase I (#79254, QIAGEN) according to the manufacturer's protocol. The concentration of RNA was measured using a spectrophotometer (NanoDrop, ND-1000). Each sample was diluted to 100 ng/µL using double distilled water.

2.3.2 cDNA synthesis

For cDNA synthesis 1 µg of RNA was reverse-transcribed in a 20 µL reaction following the manufacturer's protocol for the GoScript Reverse Transcription System (#A5003, Promega).

2.3.3 Quantitative PCR

qPCR reactions were in a 10µL volume and contained 5 µL of SYBR master mix (#B0701, Eurogentec), 4 µL of cDNA and 1 µL of 10µM primers (Table 2-3). Reactions were prepared in white 384-well plates (#04729749001, Roche) and measured using a LightCycler 480 (#05015278001, Roche) and the following parameters: 95°C for 3 min, 40 cycles of 95°C for 10 s, 60°C for 20s, 72°C for 30 s followed by a dissociation curve (66-95°C) at the end of each

run. Two technical replicates were performed for each sample and 3-4 biological replicates were used for each experiment.

2.3.4 Table of primers for qPCR

Table 2-3. Primers used for qPCR of tobacco *rbcS* genes

Gene	Sequence	Amplicon (bp)
rbcS-S1	GATACTATGATGGCAGATACTGGAC TTCAAACAAACTGCCCCTAAA	250
rbcS-S2	TCGAGACTGAGCACGGATT rbcS_S2 F1	305
rbcS-S3	TCGAGACTGAGCACGGATT rbcS_S3 F1	297
rbcS-S4	TCGAGACTGAGCACGGATT rbcS_S4 R1	293
rbcS-S5	TCGAGACTGAGCACGGATT rbcS_S5 R2	312
rbcS-T1	CTATGACGGCAGATACTGGAC AAATTAAAACAACACAACCCCTAAA	250
rbcS-T2	TCGAGACTGAGCACGGATT rbcS_T2 R2	245
rbcS-T4	TCGAGACTGAGCACGGATT GGGGAAAAACACAAGGAGAA	339
rbcS-T5	TCGAGACTGAGCACGGATT rbcS_T5 R1	246
rbcL	TTACAAAGGGCGATGCTACC CAGGGCTTTGAACCCAAATA	157
L25	CCCCTCACCACAGAGTCTGC AAGGGTGTTGTTGTCCTCAATCTT	65

2.4 Cloning

2.4.1 Golden Gate Assembly

Constructs were generated following the Golden Gate (MoClo) assembly standard (Engler et al., 2014). Enzymes *Eco31I (BsaI) (*10 U/µL) (NEB), *Bpi*I (*Bbsl*) (10 U/µL) and T4 DNA ligase (5 U/µL) were purchased from Thermo Fisher Scientific. Each reaction contained *ca.* 100 ng of each vector or PCR product. To assemble Level 0, Level 1 and Level M vectors a 20 µL reaction was set up containing the inserts and acceptor vectors (100 ng each), 10 U *Bpi*I, 2 µL Buffer G (ThermoFisher), 400 U T4 DNA ligase and 2 µL 10 mM ATP. To assemble Level -1 and Level 1 vectors 20 U *Bsa*I was used in 2 µL CutSmart Buffer (NEB) instead of *Bpi*I/Buffer G.

2.4.2 E. coli transformation

1.5 µL of the assembly mixture was used to transform 50 µL of chemically competent *E. coli* TOP10 cells. The transformation mixture was gently mixed and incubated on ice for at least 5 minutes before transfer to a heat block (42°C) for 1 minute. Cells were mixed with 250 µL of Luria broth and incubated at 37 °C for 1 hour. Positive colonies were selected by blue-white selection on IPTG (0.5mM) and X-Gal (40 µg/ml) containing LB plates with either carbenicillin, spectinomycin, or kanamycin.

2.4.3 Plasmid extraction

Colonies were used to grow 5 mL overnight cultures in LB with the appropriate antibiotic. Up to 3 mL of culture was pelleted by spinning at 17000 rcf for 5 mins and plasmid was extracted using a GeneJET Plasmid Miniprep Kit (#K0502, ThermoFisher Scientific) according to the manufacturer's instructions.

2. 5 Vectors

2.5.1 Plasmid sources and verification

Constructs were made using vectors from the GoldenGate vector and parts toolkits (#1000000044 and #1000000047, Addgene) and RGEN, gRNA parts, and associated promoters gifted by Nicola Patron (Earlham Institute, UK) (Raitskin et al. 2019). New Level 0 parts and Level 1 gRNA assemblies were verified by PCR and sequencing with the appropriate primers (Table 2-7). Other Level 1 vectors, Level M and Level 2 assemblies were verified by PCR with the appropriate primers and restriction digest.

2.5.2 *gRNA*-Cas9 vectors

The Cas-Designer tool (www.rgenome.net) (Bae et al., 2014) was used to search for 20-nt target sites adjacent to a 5'-NGG-3' motif in the genomic DNA sequences of *rbcS-T1* and *rbcS-S1*. Potential off-target sites were searched using the Cas-OFFinder tool (www.rgenome.net) (Bae et al., 2014) and two gRNA sequence (gRNA-1-1 and gRNA1-2) were selected. Full-length gRNA sequences were synthesised by PCR using a Level 0 vector containing the gRNA scaffold as a template (pICSL90010) and the target sequence included in the forward primer (Table 2-6). The PCR product was cloned directly into a Level 1 position 4 or 5 vectors (pICH47735 or pICH47736) with the Arabidopsis *U6* polymerase III promoter (pICSL90002) resulting in gRNA-1-1 and gRNA1-2.

The plant codon optimised *SpCas9* fused to a nuclear localisation signal (SV40) (pICSL90004) was expressed from the Arabidopsis *Ubiquitin 10* (P_{UBQ10}) promoter (pICSL12015) and terminated by an in-house generated terminator from the Arabidopsis *Heat Shock Protein 18.2* (HSP) (Nagaya et al., 2010) into a Level 1 position 2 binary vector (pICH47734) resulting in Level 1 *SpCas9*. For transient expression, a C-terminal fluorescent tag (*e-YFP*) was fused to *SpCas9* (pICSL90005) (*SpCas9-YFP*) and the Level 1 *SpCas9-YFP* cassette was cloned into a Level M vector (pICH8031) with gRNA1-1 and gRNA1-2 resulting in *SSU-T.* For stable expression, Level 1 *SpCas9*, gRNA1-1, gRNA1-2 and a kanamycin

resistance cassette (*nptII*) (pICSL11024) were cloned into a Level 1 position 2 binary vector (pICH47734) resulting in *SSU-1*.

2.5.3 Heterologous SSU vectors

Mature coding sequences for *AtrbcS1A* (AT1G67090) and *CrrbcS2* (X04472.1) with the *AtrbcS1A* transit peptide were obtained from a previous study (Atkinson et al. 2017). Heterologous SSU coding sequences for *HarbcS, SbrbcS* and *LgrbcS* were obtained by reverse translating mature SSU protein sequences using EMBOSS Backtranseq (EMBL-EBI). Coding sequences were synthesised as gBlock Gene Fragments that were used as templates for PCR (Table 2-7). PCR products were cloned into Level 0 CDS1 (pICH41308) and Level 0 CDS1 with no stop codon (pAGM1287) vectors.

Level 0 vectors containing *HarbcS, SbrbcS* and *LgrbcS* were assembled with the *AtrbcS1A* (AT1G67090) transit peptide that included 25 aa of the mature *AtrbcS1A* coding sequence, *AtrbcS1A* promoter and *HSP* terminator into a Level 1 position 2 vector (pICH47742) resulting in Level 1 *HA, SB* and *LG*. The Level 0 vector containing *AtrbcS1A* was assembled with the *AtrbcS1A* promoter and *HSP* terminator into a Level 1 position 2 vector (pICH47742) resulting in *1A.* The Level 0 vector containing *CrrbcS2* was assembled with the *S. lycopersicum rbcS2* promoter (X05983.1) (pICH71292) and *HSP* terminator into a Level 1 position 2 vector (pICH47742) resulting in *CR2*. For transient expression a C-terminal GFP tag (pICSL50008) was also cloned into the Level 1 vectors resulting in *1A-GFP, HA-GFP, SB-GFP, LG-GFP* and *CR-GFP*. A Level 1 hygromycin resistance cassette was generated by assembling the *hptII* coding sequence (pICSL80036) with the *A. tumefaciens* nos promoter (pICH42211) and terminator (pICH41421) resulting in Level 1 *hptII*.

For Arabidopsis transformation, Level 1 *1A, HA, SB* and *LG* vectors were cloned with a FAST-RED selectable marker (pICSL11015) into a Level 1 position 2 binary vector (pICH47734) resulting in L2-*1A,* L2-*HA,* L2-*SB* and L2-*LG*. For tobacco transformation, *CR* was assembled with *hptII* resulting in L2-*CR2*.

2.5.4 Luciferase assay vectors

Coding sequences for the NanoLuciferase (KY776554) (NLuc) and *Photinus pyralis* luciferase (firefly) (FLuc) (M15077) were gifted as Level 1 CDS1 vectors from Naomi Nakayama (University of Edinburgh). FLuc was assembled with P_{UBQ10} (pICSL12015) and the *Nos* terminator into a Level 1 position 1 vector (pICH47732) resulting in *Ubi-FLuc*. NLuc was assembled with either the *AtrbcS1A* promoter, *AtrbcS3B* promoter (AT5g38410) (pICH45244), *S. lycopersicum rbcS2* promoter (pICH71292) or *A. tumefaciens Nos* promoter (pICH42211) and the *HSP* terminator into a Level 1 position 2 vector (pICH47734) resulting in *1A-NLuc*, *3B-NLuc*, *Sl-NLuc* and *Nos-NLuc*. Level 1 *Ubi-FLuc* was assembled with either L2_*1A-NLuc*, L2_*3B-NLuc*, L2_*Sl-NLuc* or L2_*Nos-NLuc* into a Level M vector (pICH8031) (Table 2-4).

2.5.5 Table of vectors for transient expression

Table 2-4. Level 1 and Level M vector used for transient expression assays

Name	Level	Description	Map
1A-GFP	1	*AtrbcS-1A* promoter with *AtrbcS1A* coding sequence and GFP c-terminal tag	—
CR2-GFP	1	*SlrbcS2* promoter with *CrrbcS2* coding sequence and GFP c-terminal tag	—
HA-GFP	1	*AtrbcS-1A* promoter with *HarbcS* coding sequence and GFP c-terminal tag	—
LG-GFP	1	*AtrbcS-1A* promoter with *LgrbcS* coding sequence and GFP c-terminal tag	—
SB-GFP	1	*AtrbcS-1A* promoter with *SbrbcS* coding sequence and GFP c-terminal tag	—
L2_1A-NLuc	M	Ubiquitin10 promoter with firefly luciferase (fLUC) coding sequence and *AtrbcS-1A* promoter with nano luciferase (nLUC) coding sequence	Appendix A Fig. 1
L2_Nos-NLuc	M	Ubiquitin10 promoter with fLUC coding sequence and *nos* promoter with nLUC coding sequence	Appendix A Fig. 2
L2_S2-NLuc	M	Ubiquitin10 promoter with fLUC coding sequence and *SlrbcS-S2* promoter with nLUC coding sequence	Appendix A Fig. 3
L2_3B-NLuc	M	Ubiquitin10 promoter with fLUC coding sequence and *AtrbcS-3B* promoter with nLUC coding sequence	Appendix A Fig. 4
SSU-T	M	*Cas9* coding sequence coding sequence with *YFP* c-terminal tag and gRNA-1 and gRNA-2	Appendix A Fig. 5

2.5.6 Table of vectors for stable expression

Table 2-5. Level 2 vectors for stable transformation of tobacco and Arabidopsis

Name	No.	Description	Map
1A	*L2.11*	*AtrbcS1A* coding sequence with *AtrbcS1A* promoter) and FAST selectable marker for Arabidopsis transformation.	Appendix A Fig. 6
HA	*L2.13*	*HarbcS* coding sequence with *rbcS1A* promoter) and FAST selectable marker for Arabidopsis transformation.	Appendix A Fig. 7
LG	*L2.14*	*LgrbcS* coding sequence with *rbcS1A* promoter) and FAST selectable marker for Arabidopsis transformation.	Appendix A Fig. 8
SB	*L2.15*	*SbrbcS* coding sequence with *rbcS1A* promoter) and FAST selectable marker for Arabidopsis transformation.	Appendix A Fig. 9
CR2	*L2.16*	*CrrbcS2* coding sequence with *rbcS1A* promoter) and hygromycin resistance for tobacco transformation.	Appendix A Fig. 10
SSU1	*L2.91*	*Cas9* and gRNA expression construct with kanamycin resistance for tobacco transformation	Appendix A Fig. 11

2.5.7 Tables of primers for cloning

Table 2-6. Primers used for cloning PCR amplification of gRNAs for Level 0 cloning. The gRNA sequence in the forward primer is underlined.

Name	Sequence
Universal R	TGTGGTCTCTAGCGAAAAAAAGCACCGACTCGGTGCCAC
SSU1-1 F	tgtggtctcaATTGCAATGTTGCTCAAGCTAACAgtttaagagctatgctggaaacag
SSU1-2 F	tgtggtctcaATTGAGGCCTGGATCCGTATCATgtttaagagctatgctggaaacag

Table 2-7. Primers used for PCR amplification of heterologous SSUs for Level 0 cloning.

Gene	Accession	Primers
HA	XM_022136941.1	TGTCCGAGAAGACACaggtATGAAAGTTTGGCCACCA
		TGTCCGAGAAGACACcgaaTAGCCGTCTGGTCGGGAT
SB	AB564718.1	CTCCACAGAAGACGTaggtATGCAAGTTTGGCCAGCT
		CTCCACAGAAGACGTcgaaTCAGATCCAGCTGGCTTA
LG	AJ786661.1	TGGCTCAGAAGACGTaggtATGCAAGTGTGGCCACCC
		TGGCTCAGAAGACGTcgaaGATGGGGGTTTGTAGGCT

Table 2-8. Primers used to verify plasmids by colony PCR and sequencing.

Name	Sequence
L0 F0017	TTGAGTGAGCTGATACCGCT
L0 R0016	GTCTCATGAGCGGATACATATTTGAATG
L1 F0229	GAACCCTGTGGTTGGCATGCACATAC
L1 R0230	CTGGTGGCAGGATATATTGTGGTG
L2 F0231	GTGGTGTAAACAAATTGACGC
L2 R0232	GGATAAACCTTTTCACGCCC

2.6 Protein methods

2.6.1 Protein extraction

Frozen plant tissue (20-100 mg) was homogenised in a TissueLyser using a two 3mm tungsten beads. Extraction buffer containing 25mM Tris-HCl pH 7.5, 1mM EDTA, 10% glycerol, 0.1% TritonX-100 (Sigma Aldrich), 150 mM NaCl, 1mM DTT and a cOmplete EDTA-free protease inhibitor cocktail (COEDTAF, Roche). Samples and buffer were vortexed and spun at 5000 rcf at 4°C for 5 mins. The supernatant was recovered and 1% (w/v) of lithium dodecyl sulphate (LDS) and 1µL β-mercaptoethanol (per 100 µL) was added before vortexing again. Samples were heated to 100 °C for 1 minute and used directly or flash frozen and stored at - 80 °C.

2.6.2 Bradford Assay

Protein extracts (10 µL) were combined with 150 µL of Pierce 660nm Protein Assay Reagent (22660, ThermoFisher) according to the manufacturer's instructions. Pierce bovine serum albumin pre-diluted standards (23208, ThermoFisher) were used to generate a standard curve (2000-250 µg/mL). Measurements at 660 nm were performed using a FLUOstar Omega plate reader (BMG Labtech).

2.6.3 SDS page

Protein extracts were prepared for SDS-PAGE by heating at 100 °C for 3 minutes in 1X Bolt LDS sample buffer (B0007, Invitrogen) and 20mM DTT. Samples were run on 12-well NuPAGE 12% Bis-Tris gels (NP0342, Invitrogen) in 1x MOPS SDS running buffer (NP0001, Invitrogen) at 150V for 70 mins alongside 5µL of SeeBlue plus 2 pre-stained protein standard (LC5925, Invitrogen).

2.6.4 Western Blotting

Denatured samples were transferred to a PVDF membrane using the iBlot2 gel transfer device (IB21001, Invitrogen) for 8 minutes (20 V for 1 minute, 23 V for 4 minutes and 25 V for 2 minutes). Membranes were blocked in 5% (v/v) semi-skimmed milk in TBST (20 mM Tris, 150 mM NaCl, 0.1% (w/v) Tween 20) for 1 hour and incubated in a primary antibody overnight. Primary antibodies used to detect the proteins of interest were rabbit serum raised against wheat Rubisco anti-Rubisco (1:10 000 dilution) (Howe et al., 1982), rabbit anti-histone H3 (ab18521, Abcam) (1:10 000 dilution) and mouse anti-CRISPR-Cas9 (ab191468, Abcam) (1:5000 dilution). Membranes were subsequently incubated with a 1:10,000 dilution of Li-Cor IRDye® 800CW goat anti-rabbit IgG (Li-Cor Inc.) or HRP-conjugated anti-mouse IgG (ab6728, Abcam). Membranes incubated with Li-Cor antibodies were viewed on a Li-Cor Odyssey CLx Imager. Membranes incubated with HRP-conjugated antibodies were incubated for 5 minutes in Pierce ECL western blotting substrate solution (32209, ThermoScientific) and exposed to an X-ray film.

2.6.5 Total protein and Rubisco content analysis

Leaf samples of 5.88 cm^2 were collected from the youngest fully expanded leaves of four-week-old plants, frozen on liquid nitrogen and stored at -80 °C before extraction. Samples were processed by Doug Orr (Lancaster University, UK) as follows. The extraction buffer contained 50 mM Bicine-NaOH pH 8.2, 20 mM $MgCl_2$, 1 mM EDTA, 2 mM benzamidine, 5 mM ε-aminocaproic acid, 50 mM β-mercaptoethanol, 10 mM dithiothriotol, 1% (v/v) protease inhibitor cocktail (Sigma-Aldrich, Mo, USA), and 1 mM phenylmethylsulphonyl fluoride. Samples were ground rapidly in an ice-cold mortar and pestle in 250 µL of extraction buffer for ca. 1 min followed by 1 min centrifugation at 4 °C, 14700 g. 90 µL of the supernatant was then mixed with 100 µL of CABP binding buffer which contained 100 mM Bicine-NaOH (pH 8.2), 20 mM $MgCl_2$, 20 mM $NaHCO_3$, 1.2 mM (37 kBq/µmol) [^{14}C]CABP (carboxyarabintol-1,5-bisphosphate), incubated at RT for 25 min, and Rubisco content determined via

[^{14}C]CABP binding (Sharwood et al., 2016). Bradford assay (Bradford, 1976) was used to determine total soluble protein in the same supernatant as prepared for Rubisco content analysis.

2.7 Plant materials and growth conditions

2.7.1 Arabidopsis lines and growth conditions

Arabidopsis wild-type (*Arabidopsis thaliana* (L.) Heynh. Col-0) the *1a3b* mutant (GABI_608F01 (At1g67090); SALK_117835 (At5g38410) and the T-DNA insertion lines were sown on a compost and sand mix (Levington F2+S) and stratified for 3-4 days at 4°C. Seeds were germinated at 22°C, ambient CO_2, 60-70% relative humidity and in a 12 hour photoperiod (200 μmol photons m^{-2} s^{-1}) unless otherwise described. Seedlings were transplanted to trays or pots 10-12 days after stratification. All seedlings were transplanted to trays 4X6 trays 10-12 days after stratification. For material for floral dipping, the *1a3b* mutant was grown in a 16-hour photoperiod. Seeds obtained from primary transformants (T_1 generation) and the progeny (T_2 generation) were grown at 25°C in a 16-hour photoperiod (170 μmol photons m^{-2} s^{-1}) in a controlled growth chamber (Percival, AR-36L3). The T_3 generation of transgenic lines was used for growth analyses, molecular analyses and photosynthetic measurements. Transgenic lines for growth experiments were grown at 22°C in a 12-hour photoperiod (170 μmol photons m^{-2} s^{-1}). All growth chambers were supplied with cool white fluorescent bulbs with top illumination unless otherwise stated.

2.7.2 Tobacco lines and growth conditions

Wild-type tobacco plants (*Nicotiana tabacum* (L.) Petite Havana) seeds were gifted by Christine Raines (University of Essex, UK). Wild-type and transgenic tobacco (T_1 generation) seeds were sown on a compost and sand mix (Levington F2+S) and germinated at 25°C, 60-

70% relative humidity in a 16-hour photoperiod (170 μmol photons m^{-2} s^{-1}) in a controlled environment growth chamber (Percival, AR-36L3) unless otherwise described.

2.7.3 Tobacco growth analysis conditions

Seedlings were germinated as described in section 2.7.2. After 14 days wild-type and three knockout lines were transplanted to 3 L pots. Plants were maintained in a greenhouse ((20-21 °C (day), 18 °C (night)) under natural light supplemented with 270-320 μmol photons m^{-2} s^{-1} in a 15 h photoperiod. Plants were rotated every other day and supplemented with Hoaglands solution weekly. For qPCR analysis of mature leaves, 60 mg of leaf tissue was collected from the youngest fully expanded leaves of three replicate plants.

2.7.4 Tobacco light experiment conditions

Seeds used in the time course and light experiments were surface sterilised (section 2.5.3) and grown on MS media (1x MS (Sigma Aldrich), 0.8% (w/v) agar) for ten days. Light grown seedlings were grown under an irradiance of 150 μmol photons m^{-2} s^{-1} in a 12 h photoperiod at 25°C (Percival, AR-36L3). Dark-grown seedlings were grown at 25°C (Percival) in a controlled environment cabinet kept in a dark room. Four replicate plates of seedlings were used for each time point and light treatment with 20-30 whole seedlings harvested for each replicate.

2.7.5 Light treatments

All cabinets were maintained at a temperature of 25°C. Ten-day-old etiolated seedlings were exposed to 150 μmol photons m^{-2} s^{-1} white light for 12 h (AR-36L3, Percival Scientific). For the light quality experiments, ten-day-old seedlings were exposed to a single pulse (90 μmol photons m^{-2} s^{-1} for 10 min) of red light (630-670 nm), far-red light (720-740 nm), red light immediately followed by the same far-red light treatment, or blue-light (440-460 nm). All

samples were harvested under a green safe light. Four replicate plates of seedlings were used for each time point and light treatment with 20-30 whole seedlings harvested for each replicate.

2.7.6 Seed sterilisation

Tobacco seeds were sterilised overnight in an Eppendorf tube within a sealed container with 50mL bleach and 1.5mL concentrated HCl.

2.8 Transformation methods

2.8.1 Agrobacterium transformation

AGL1 and GV101 *A. tumefacians* strains were transformed by mixing 50 µL of electrocompetent cells with *ca.* 100 ng of plasmid DNA. After gentle mixing the cells were transferred to an electroporation cuvette (FBR-102, Scientific Laboratory Supplies) and electroporated at 1800 mV in an electroporator (2510, Eppendorf). Cells were recovered in 1 mL of SOC buffer and incubated in a shaking incubator at 28 °C for four hours. 10 µL of cells were plated on LB containing the appropriate antibiotics. Colonies were verified by colony PCR with insert specific primers (Table 2-8).

2.8.2 Floral dipping

Verified *Agrobacterium* colonies were used to grow 5 mL LB cultures (with antibiotics) at 28°C for 24 H. Overnight cultures were used to inoculate 200 mL of LB (with antibiotics) and 150 µM acetosyringone and grown at 28 °C for 24 H. Cells were harvested by centrifugation (5000 rpm for 15 minutes at 4°C) and resuspended in 200 mL infiltration medium (0.5x MS, 5% sucrose, 3mM MES, 0.1% Silwet (v/v), 150 µM acetosyringone). Six-week-old plants were dipped in the suspension for 2 minutes and again five days later. T_1 seeds were selected by FAST-red fluorescence using a green light 530 nm LED torch (Joyland) viewed through a red filter.

2.8.3 Agroinfiltration

Verified *Agrobacterium* colonies were used to grow 5 mL cultures as described in section 2.8.2. 5 mL cultures were used to inoculate 15 mL LB cultures (with antibiotics) were spun down at 5000 rcf for 10 minutes, resuspended in 10mM $MgCl_2$ and diluted to a final OD_{600} of 0.8. Diluted cultures were syringe infiltrated into the youngest fully expanded leaves of three- to four- week old tobacco plants. Leaf discs were harvested 48 hours post-infiltration and the expression of *SpCas9-YFP* confirmed by confocal laser scanning microscopy (Leica TCS SP8).

2.8.4 Tissue culture (from sterile plants)

Sterile tobacco seeds (section 2.5.3) were grown in Magenta™ GA-7 boxes (V8505; Sigma Aldrich) on non-selective media (NSM) (1x Murashige and Skoog Basal Salt Mixture (Sigma Aldrich), 3% sucrose (VWR), 0.8% agar (Sigma Aldrich), pH 5.7). Plants were grown under an irradiance of 90 μmol photons m^{-2} s^{-1} in a 16-hour photoperiod at 21°C. A 150 mL *Agrobacterium* (AGL1) culture was prepared as described in section 2.8.2. Cultures were grown for 24 hours, spun at 5000 rcf for 10 minutes and resuspended in the same volume of liquid MS (Sigma Aldrich). All of the following procedures were performed in a laminar flow cabinet with autoclaved reagents and materials. Leaves from six-week-old plants were cut into 2 cm^2 pieces and incubated for 30 minutes in the AGL1 suspension before transfer to NSM containing 0.1 mg/L indole-3-butyric acid (IBA) (Sigma Aldrich, 57310) and 1 mg/L 6-benzylaminopurine (Sigma Aldrich, B3408). After 2 days, explants were washed in liquid 1 x MS to remove *A. tumefacians* and transferred to a selective media (500 mg/L augmentin and appropriate antibiotic for selection). Kanamycin (100 mg/L) or hygromycin (30 mg/L) or a combination of the two antibiotics were used to select transformants. Shoots were excised from explants four to five weeks after incubation and transferred to a rooting media in

Magenta™ GA-7 vessels. Regenerated plants were allowed to establish on soil after the formation of lateral roots.

2.8.5 Tissue culture (from infiltrated leaves)

Leaves were infiltrated with a vector (*SSU-T*) as described in section *2.8.3.* Leaves were removed three days after transfection and surface sterilised by immersion in 5% bleach for 5 minutes and washed three times with sterile double distilled water. Explants were made as described in *2.6.4* and cultured on NSM containing augmentin (500 mg/L).

2.9 Protoplast methods

2.9.1 Media compositions

MGG (pH 5.6)

Component	Concentration (g/L)
Glucose	45
Glycine	25
MES	0.7
Onozuka R10	2
Maceroenzyme	0.6
Driselase	0.8

MMM

Component	Concentration
Mannitol	0.5 M
$MgCl_2$	15 mM
MES	0.1% (w/v)

PEG (pH 8)

Component	Concentration
Mannitol	0.4 M
$Ca(NO_3)_2.2H_2O$	0.1 M
PEG4000	40% (w/v)

2.9.2 Protoplast extraction

Mesophyll protoplasts were harvested from three-week old sterile tobacco leaves by enzymatic digestion in 10 mL of filter sterilised MGG media for three hours (Chupeau et al., 2013). All centrifugation steps were performed at 70 rcf for 5 minutes. The protoplast solution was applied to a 100 μm cell strainer (Thermo Fisher) and protoplasts were collected in a 50 mL falcon tube. After washing three times in 3 mL MGG (without enzymes) protoplasts were resuspended in 3 mL MMM and the concentration of cells measured by applying 10 μL of protoplast solution to a haemocytometer.

2.9.3 PEG transformation

Plasmids were purified from 10 mL TOP10 cultures using a GeneJET™ Endo-free Plasmid Maxiprep Kit (Thermo Scientific™). For transformation 300 μL of protoplasts (6 x 10^4 protoplasts/reaction) were gently mixed with an equal volume of PEG solution and 10 μg of plasmid. After one minute the PEG solution was diluted with 1.4 mL of MGG and protoplasts were incubated in the dark for 1 H. Protoplasts were pelleted, resuspended in 1 mL of MGG and incubated in the dark for 24 H at room temperature.

2.9.4 Dual luciferase assay

Protoplasts were concentrated by centrifugation and resuspended in 100 μL of MGG (without enzymes). Luciferase activity was measured using he Nano-Glo Dual-Luciferase Reporter Assay System (Promega) according to the manufacturer's instructions. In brief, 40 μL of

protoplasts were mixed with an equal volume of ONE-Glo EX luciferase assay reagent and firefly luminescence was measured using a FLUOstar Omega plate reader (BMG Labtech). After the measurement 40 µL of NanoDLR Stop & Glo reagent was added and the resulting NanoLuc activity was measured. The relative fluorescence data was expressed as a ratio of the NanoLuc fluorescence divided by the firefly luminescence.

2.10 Photosynthetic measurements

2.10.1 Chlorophyll extraction

Chlorophyll was extracted on a leaf area basis (58.9 mm^2) for all samples in an acetone solution containing 10mM Tris-HCl (pH 8.0) and 80% (v/v) acetone. Leaf discs were homogenised as described in section 2.1.1, resuspended in 1 mL of acetone solution and spun down at 14000 rpm for 5 minutes. 950 µL of supernatant was transferred to a quartz cuvette and absorbance at 663 nm and 646 nm was measured using a spectrophotometer. Chlorophyll *a* and chlorophyll *b* were calculated according to Porra et al., 1989.

2.10.2 Chlorophyll fluorescence

Dark-adapted leaves were to determine the maximum quantum yield of photosystem II (F_v/F_m) using a Hansatech Handy PEA chlorophyll fluorimeter (Hansatech Instruments). F_v/F_m measurements were taken on the final day of growth experiments prior to harvesting.

2.10.3 Gas exchange and estimation of photosynthetic parameters

The response of photosynthesis (A) to irradiance (Q) and CO_2 concentration (C_i) was measured using a portable gas exchange system (LI-COR 6400-XT, LICOR). The A/C_i response was measured under saturating light intensity (1500 µmol photons m^{-2} s^{-1}) of light. The first measurement was taken at an external CO_2 concentration (C_a) of 400 µmol mol^{-1} and decreased to 50 µmol mol^{-1} in increments of 50. The upper part of the *A/Ci* response was measured from

500-2000 μmol mol^{-1} in increments of 200. The *A/Q* response was performed at atmospheric C_a (400 μmol mol^{-1}). All measurements were taken at a leaf temperature (T_{leaf}) of 25 °C and relative humidity of 60-70% was maintained. The maximum carboxylation rate ($V_{c,max}$) of Rubisco and maximum electron transport rate (J_{max}) were estimated by modelling the A/C_i response to the biochemical models of C3 photosynthesis (Ethier and Livingston, 2004; Farquhar et al., 1980). The maximum rate of photosynthesis (A_{max}) was estimated from the *A/Q* response following Marshall and Biscoe (1980) and Monteith (1991).

2.11 Growth analysis measurements

2.11.1 Leaf area and height measurements

Arabidopsis plants were photographed at the same time for 22-days and the rosette area quantified using iDIEL Plant software (Dobrescu et al., 2017). The leaves and stems of 45-day old tobacco plants were harvested and photographed separately. Total leaf area and leaf count were measured using iDIEL Plant software. Plant height was determined by measuring the stem length using ImageJ software (Schneider et al., 2012).

2.11.2 Fresh and dry weight measurements

Whole Arabidopsis rosettes were harvested after 34 days and tobacco leaves and stems harvested as described in section 2.11.1. Fresh weight was measured immediately, and dry weight was measured after incubation at 80 °C for one week.

2.12 Statistical analyses

Statistical analyses were performed using Prism software (Graphpad). A *t*-test was used to determine if the means of two groups of normally distributed data were significantly different from each other. A one-way ANOVA was used to identify significant differences among

normally distributed data from more than two groups. Groups that were significantly different as determined by a one-way ANOVA were identified using Tukey's HSD test. The *P* values for each statistical test are reported in figure legends.

Chapter 3

Characterising the *rbcS* family in *Nicotiana tabacum*

3.1 Introduction

Nicotiana tabacum (tobacco) is an important non-food crop species and platform for expressing proteins, including biopharmaceuticals (Paul and Ma, 2016). Foreign genes can be efficiently expressed from the nuclear and chloroplast genomes and the regeneration of transgenic tobacco plants is well established. Therefore, tobacco is a useful model for engineering chloroplast proteins including Rubisco (Kanevski et al., 1999; Whitney et al., 1999; Whitney and Sharwood, 2008). However, it has been challenging to characterise the SSU encoding *rbcS* family because of the genomic complexity of tobacco and a lack of bioinformatics resources. Understanding the contribution of individual *rbcS* isoforms to Rubisco in tobacco will be important for developing tobacco further as an *in planta* platform to manipulate the LSU and SSUs simultaneously.

The polyploid genome of tobacco ($2n = 4x = 48$) arose relatively recently (*ca.* 0.2 Mya) from a cross between *Nicotiana sylvestris* (maternal, S-genome donor) and *Nicotiana tomentosiformis* (paternal, T-genome donor) (Clarkson et al. 2005; Leitch et al. 2008). As well as a large genome size (4.5 Gb), the majority of the tobacco genome is composed of repeats (*ca.* 75%) (Edwards et al. 2017). It is difficult to assemble the genomes of polyploid species, particularly for tobacco because of substantial genetic complexity (Schatz et al., 2012). Currently, more than 1,900 varieties of tobacco are cultivated commercially and significant genetic variation exists between different cultivars (USDA, 2019). For example, one of the most widely cultivated varieties, Burley, has a chlorophyll deficiency that is conferred by a recessive genotype (Legg et al., 1977). Next-generation sequencing (NGS) of three

commercially grown species facilitated *de novo* assemblies that covered *ca.* 80% of the genomes; however, the assemblies were highly fragmented (Sierro et al. 2014). Gaps in incomplete scaffolds (denoted by a string of N's) often arise between repetitive regions, homologous genes, or gene families (Denton et al. 2014). Recently, an improved genome assembly was published with better coverage (90%) and included a chromosome map (Edwards et al. 2017). Further improvements in coverage of the gaps in the tobacco genome assemblies will assist functional genomic studies in addition to gene editing approaches.

Tobacco has at least thirteen *rbcS* isoforms, with six originating from the *N. sylvestris* genome (referred to here as *rbcS-S*) and seven from the *N. tomentosiformis* genome (referred to here as *rbcS-T*) (Gong et al., 2014). Initially, three unique isoforms were amplified from genomic DNA isolated from tobacco leaves (Mazur and Chui, 1985; O'Neal et al., 1987). One isoform was a pseudogene that produced a truncated peptide (O'Neal et al., 1987) and the remaining two isoforms had a novel gene structure compared to *rbcS* isoforms from most other land plants. The vast majority of *rbcS* isoforms in higher plants have two introns (Dean et al., 1989). However, two isoforms from tobacco and some family members isolated from other *Solanaceae* species (e.g. potato, petunia) have three introns (Mazur and Chui, 1985; Dean et al., 1989). Following those early characterisation studies, the remaining 10 isoforms were partially sequenced by Gong et al. (2014). Often *rbcS* isoforms occur as tandem duplicates, which are defined as genes belonging to the same family and located within 100 kB of one another (Hanada et al. 2008). For example, three out of four *rbcS* isoforms in Arabidopsis are a result of tandem duplication and are located within an 8 kb region. Whether tandem duplication has played a key role in the evolution of the tobacco *rbcS* family remains unclear as the precise chromosomal locations and full-length sequences of the members of the tobacco *rbcS* family have not yet been described. Furthermore, the contribution of individual *rbcS* isoforms to tobacco Rubisco is unknown.

Rubisco small subunit genes are known to be expressed differentially during development (Suzuki et al. 2009; Wanner and Gruissem 1991) and environmental conditions, including temperature (Yoon et al. 2001), light (Dedonder et al. 1993; Sawchuk et al. 2008), and CO_2 concentration (Cheng et al., 1998). Differential regulation also occurs in dark-grown seedlings. For example, three out of four *rbcS* isoforms and three out of five *rbcS* isoforms are expressed in dark-grown Arabidopsis and tomato seedlings, respectively, albeit at low levels (Dedonder et al. 1993; Sugita and Gruissem, 1987). Distinct photoreceptors likely mediate the light-induced transcription of *rbcS* genes by perceiving changes in the ratio of red light to far-red light (RL:FRL) and blue light to green light (BL:GL) light (Dedonder et al. 1993). For example, small families of phytochromes regulate responses to RL:FRL by converting between RL-absorbing (P_{fr}) and far-red absorbing forms (P_r). The transcription of three out of four *rbcS* isoforms in Arabidopsis is phytochrome-dependent, while one isoform, *rbcS-1B*, is regulated in a phytochrome-independent manner (Dedonder et al., 1993). In Arabidopsis the *rbcS* isoforms are differentially regulated by BL (Dedonder et al., 1993; Wehmeyer et al., 1990) likely via families of cryptochrome (Cashmore et al., 1999) and phototropin (Briggs et al., 2001). RL, FRL, and BL increases the accumulation of total *rbcS* mRNA transcripts in tobacco (Wehmeyer et al., 1990). However, the expression patterns of individual tobacco *rbcS* isoforms in response to different wavelengths of light is unknown.

In this chapter, the individual *rbcS* isoforms in tobacco were characterised to identify the predominant isoforms contributing to Rubisco in tobacco leaves. Full-length genomic sequences for eight out of ten isoforms were amplified from tobacco (cv. Petite Havana) and mapped to unique regions on the tobacco genome. Transcripts for two isoforms, *rbcS-S1* and *rbcS*-T1, were found to be the most abundant *rbcS* transcripts in mature plants and seedlings. The abundance of *rbcS* transcripts in response to irradiance and different wavelengths of light was measured to establish the differential expression of individual *rbcS* isoforms in tobacco. This work has provided the first insight of differential expression of the tobacco *rbcS* family.

3.2 Results

3.2.1 Identification of tobacco rbcS isoforms

The genomic DNA partial-coding sequences of thirteen tobacco *rbcS* isoforms (Gong et al., 2014) were used to BLAST search the tobacco genome assembly on the Sol Genomics database. Nine out of thirteen isoforms mapped to unique regions on the tobacco TN90 assembly (Sierro et al., 2014) (Table 3-1). In contrast, the two isoforms of *rbcS-S1* (*rbcS-S1a* and *rbcS-S1b*) and *rbcS-T4* (*rbcS-T4a and rbcS-T4b*) could not be mapped to unique loci (i.e. they mapped to the same locations). The partial-coding sequences of *rbcS-S1a* and *rbcS-S1b* shared a 98.5% nucleotide identity, while *rbcS-T4a* and *rbcS-T4b* were also nearly identical (98.1% nucleotide identity) (Fig. 3-2). As it was not possible to discriminate between the genomic locations of *a* and *b* homologs, the two pairs of homologous isoforms were referred to here as *rbcS-S1* and *rbcS-T4*. The isoforms *rbcS-T3a* and *rbcS-T3b* had a lower sequence similarity (97%) compared to *rbcS-S1* and *rbcS-T4*. However, no significant matches were found between *rbcS-T3b* and the available genome assembly. Furthermore, *rbcS-T3a* could not be aligned to a unique locus as it is nearly identical (99%) to *rbcS-T2*. One locus matched significantly to but was not identical to *rbcS-T3a* or *rbcS-T2*. Therefore, both isoforms will subsequently be described as *rbcS-T2/3*. The partial-coding sequences of all thirteen isoforms were later used to BLAST search a second genome assembly following the publication of the first genetic map for tobacco (Edwards et al., 2017). Two isoforms, *rbcS-S1* and *rbcS-T1*, were located on unique chromosomes (chromosome 21 and chromosome 14, respectively) (Table 3-1). The remaining isoforms for both the maternal (*rbcS-S*) and paternal (*rbcS-T*) isoforms were located in tandem on separate chromosomes. Specifically, the four *rbcS-T* isoforms *rbcS-T2*, *rbcS-T3a/*b, *rbcS-T4* and *rbcS-T5* were located on chromosome 17 within a region of approximately 188 Kb in size. Similarly, the four *rbcS-S* isoforms *rbcS-S2*, *rbcS-S3*, *rbcS-S4* and *rbcS-S5* were located within a *ca.* 192 Kb region on chromosome 3.

Gene-specific primers for nine isoforms (*rbcS-S1*, *rbcS-S2*, *rbcS-S3, rbcS-S4, rbcS-S5, rbcS-T1, rbcS-T2/3, rbcS-T4,* and *rbcS-T5*) were designed based on the initial draft genome assembly (Sierro et al., 2014) (Table 2-1). Primer binding sites in the 3' and 5' untranslated regions (UTRs) were chosen to amplify full-length coding sequences (Appendix B Fig. 1). A single PCR product was obtained for eight of the isoforms, but no band was amplified for *rbcS-T2* (Fig. 3–1A). Two additional pairs of primers were used to attempt to amplify from the remaining unique regions within the available *rbcS-T2* sequence, but neither primer pair yielded a PCR product. Unfortunately, the sequences in the scaffold surrounding *rbcS-T2* were ambiguous (denoted by a string of 'N's), which prevented the design of further primers (i.e., from the upstream or downstream regions of the coding sequence). The full-length coding sequences for the eight amplified isoforms were obtained by Sanger sequencing (Appendix B Fig. 1). Exon-intron junctions were predicted from the partial-coding sequences (Gong et al., 2014) and confirmed by PCR and sequencing using a cDNA template (Fig. 3–1B) (Appendix B Fig. 2). Six out of eight isoforms (*rbcS-S2*, *rbcS-S3, rbcS-S4, rbcS-S5, rbcS- rbcS-T4,* and *rbcS*-T5) had two introns in the full-length sequences. As reported previously, the two isoforms *rbcS-S1* and *rbcS-T1* had an additional intron that occurred in the last exon (Gong et al., 2014). *rbcS-S1* and *rbcS-T1* each encoded a 543-bp coding sequence (180 amino acids) and the remaining six sequenced isoforms each encoded a 546-bp coding sequence (181 amino acids).

Table 3-1. Genomic location of thirteen Rubisco small subunit genes (rbcS) in tobacco (Nicotiana tabacum). Thirteen partial-coding sequences (Gong et al., 2014) were used to BLAST search the tobacco genome assemblies on the Sol Genomics database.

No.	*Isoform*	*Scaffold* [a]	*Location (bp)*[a]	*Chromosome* [b]	*Location* [b]
1	*rbcS-S1a*	SS1336	810318-811169	Nt21	11725909-11726259
2	*rbcS-S1b*	Maps to same region as *rbcS-S1a*	-	-	-
3	*rbcS-S2*	SS4468	754873-755617	Nt03	46963643-46964387
4	*rbcS-S3*	SS4468	554204-554937	Nt03	46873734-46874467
5	*rbcS-S4*	SS4468	399989-400743	Nt03	46911953-46912707
6	*rbcS-S5*	SS4468	456463-457081	Nt03	46771193-46772295
7	*rbcS-T1*	SS2179	301404-302119	Nt14	90863242-90863574
8	*rbcS-T2*	SS17012	102405-102957	Nt17	208193244-208193991
9	*rbcS-T3a*	Maps to same region as *rbcS-T2*	-	Nt17	208121436-208122184
10	*rbcS-T3b*	Maps to same region as *rbcS-T2*	-	Nt17	Maps to same region as *rbcS-T3a*
11	*rbcS-T4a*	SS17012	88923-89705	Nt17	208180177-208180959
12	*rbcS-T4b*	Maps to same region as *rbcS-T4a*			
13	*rbcS-T5*	SS17012	138156-139280	Nt17	208193500-208193991

[a] Scaffold and location from the TN90 genome assembly (Sierro et al., 2014)

[b] Chromosome and location from Edwards genome assembly (Edwards et al., 2017)

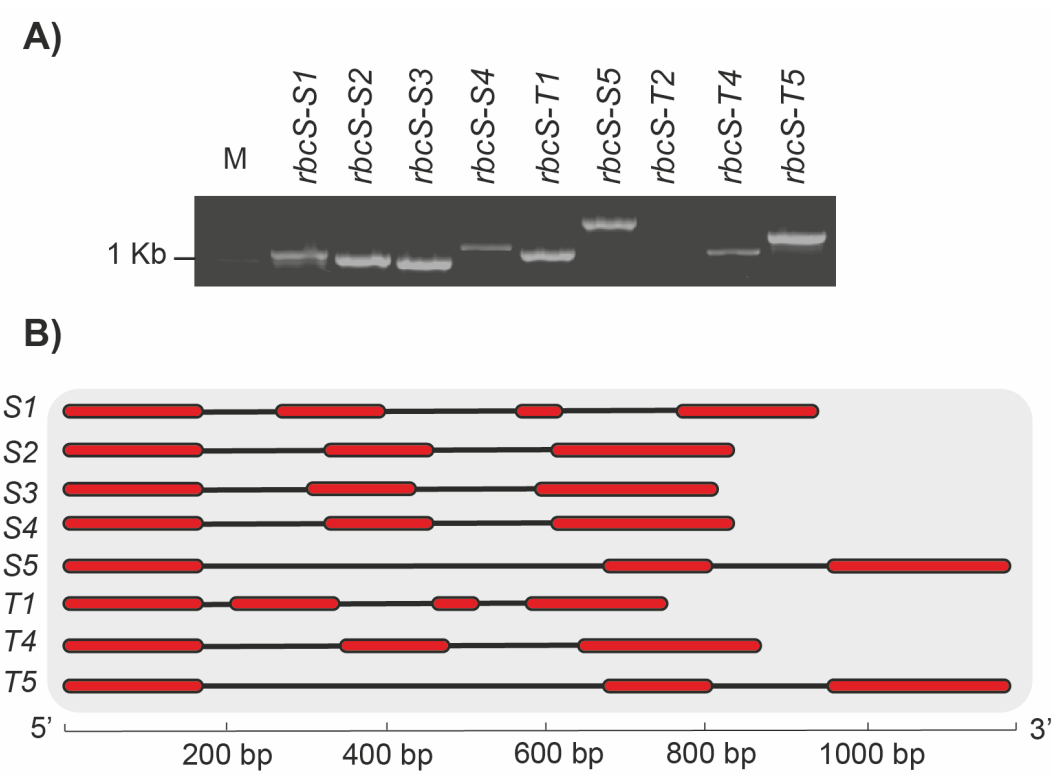


Figure 3-1. Genomic DNA sequences of eight Rubisco small subunit gene (*rbcS*) isoforms in tobacco. (A) Gene-specific primers (Table 2–1; F1 and R2) amplify eight out of nine *rbcS* isoforms from tobacco DNA. **(B)** Two isoforms are encoded by four exons (*rbcS-S1* and *rbcS-T1*) and the remaining isoforms are encoded by three exons. Exon-exon junctions are at conserved positions while the intron length varies between isoforms. Exon-intron boundaries confirmed by Sanger sequencing of PCR and RT-PCR products with gene-specific primers. The full-length sequences and primer binding sites are shown in Appendix B Fig. 1.

A multiple sequence alignment was used to determine the percent identity between the eight amplified full-length nucleotide sequences (Fig. 3–2A, B) (Appendix B Fig. 2). The coding regions were highly conserved with nucleotide identities ranging from *ca.* 86% to 96.7 %. The unlinked isoforms, *rbcS-S1* and *rbcS-T1,* each had the lowest percent identity to the remaining isoforms but were 94.7% identical to each other. All the remaining *rbcS-S* and *rbcS-T* isoforms clustered on chromosome 3 and 17, respectively, were > 92% identical overall. The coding-sequences of the nine isoforms translated into seven unique amino acid sequences, which were highly conserved. Most of the amino acid substitutions occurred in the chloroplast transit

peptide, which was predicted to consist of 57 or 58 amino acids (Fig. 3–3). The transit peptide encoded by *rbcS-T4* was particularly divergent because of a premature stop codon at position 56. The amino acid sequences of the mature SSU peptide were more conserved and had percent identities ranging from 93% to 100% (Fig. 3–4, 5). The mature peptides of *rbcS-S2* and *rbcS-S3* were identical while the remaining isoforms encoded unique peptides. *rbcS-T1* and *rbcS-T4* had unique amino acid substitutions that did not occur in the remaining SSUs. *rbcS-T4* had a single substitution (W70R) that occurred in the βb sheet. rbcST1 was the most divergent with six unique substitutions, including one in the βa-βb loop (H48R).

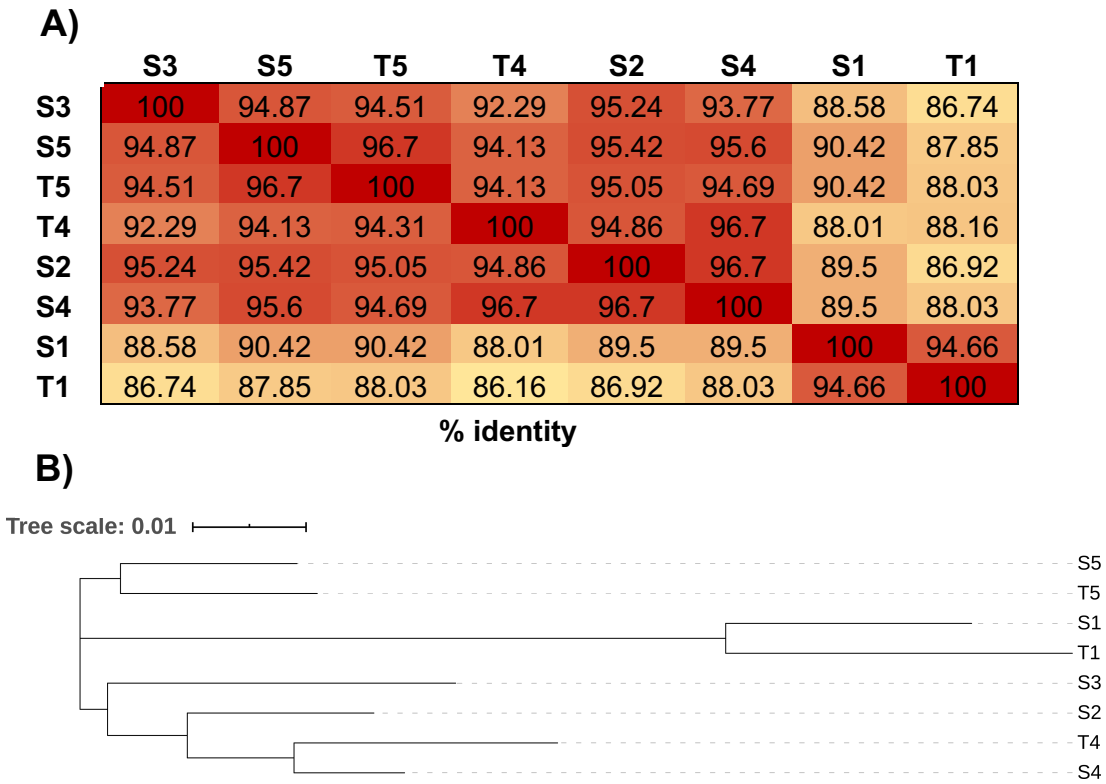


Figure 3-2. Coding sequence similarity among eight tobacco *rbcS* isoforms. (A) Coding sequences were obtained by Sanger sequencing of RT-PCR products amplified with gene-specific primers. Percent identity was determined by multiple sequence alignment with Clustal Omega (EMBL-EBI) (Appendix B Fig. 2). **(B)** Phylogenetic tree produced from the multiple sequence alignment shows the evolutionary relationships between the coding sequences.

A)

	S1	T1	S3	T4	S5	T5	S2	S4
S1	100	98.25	80.7	85.71	87.72	87.72	85.96	84.21
T1	98.25	100	82.46	85.71	87.72	87.72	85.96	84.21
S3	80.7	82.46	100	85.96	91.38	91.38	87.93	84.48
T4	85.71	85.71	85.96	100	94.74	94.74	92.98	92.98
S5	87.72	87.72	91.38	94.74	100	100	96.55	93.1
T5	87.72	87.72	91.38	94.74	100	100	96.55	93.1
S2	85.96	85.96	87.93	92.98	96.55	96.55	100	96.55
S4	84.21	84.21	84.48	92.98	93.1	93.1	96.55	100

B)

```
S1      MASSVL-SSAAVATRSNVAQANMVAPFTGLKSAASFPVSRKQNLDITSIASNGGRVQC   57
T1      MASSVL-SSAAVATRTNVAQANMVAPFTGLKSAASFPVSRKQNLDITSIASNGGRVQC   57
S3      MAFLIMSSAAAVATGTNAAQASMIAPFTGLKSATSFPVSRKQNLDITSIASNGGRVQC   58
T4      MASSVISSAAAVATGANAAQASMVAPFTGLKSAYSFPVSRKQNLDITSIASNGGRV*C   57
S5      MASSVMSSAAAVATGANAAQASMVAPFTGLKSATSFPVSRKQNLDITSIASNGGRVQC   58
T5      MASSVMSSAAAVATGANAAQASMVAPFTGLKSATSFPVSRKQNLDITSIASNGGRVQC   58
S2      MASSVMSSAAAVATGANAAQASMVAPFTGLKSASSFPVTRKQNLDITSIASNGGRVQC   58
S4      MASSVISSAVAVATGANAAQASMVAPFTGLKSASSFPVTRKQNLDITSIASNGGRVQC   58
        **  :: *:.**** :*.***.*:********* ****:***************** *
```

Figure 3-3. Rubisco small subunit transit peptides encoded by eight *rbcS* isoforms in tobacco. (A) Percent similarity of transit peptide sequences determined by multiple sequence alignment with Clustal Omega (EMBL-EBI). **(B)** Multiple sequence alignment of transit peptide sequences derived from gene-specific PCR of cDNA sequences. Transit peptides are formed of 57 or 58 amino acid residues, conserved residues indicated by asterisk and partially-conserved residues indicated by a semi-colon or full stop.

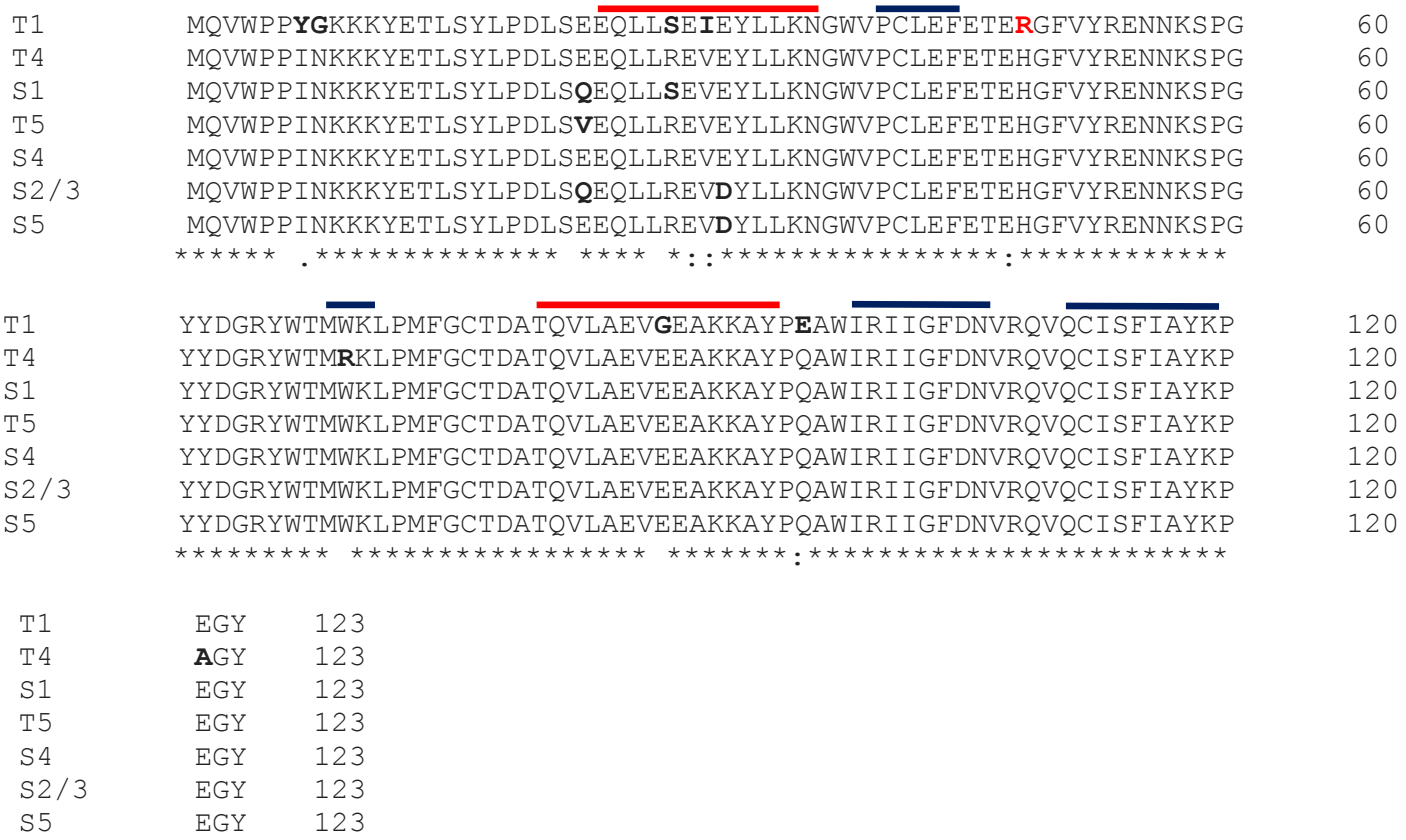


Figure 3-4. Multiple sequence alignment of seven mature SSU sequences from tobacco. Two isoforms (*rbcS-S2* and *rbcS-S3*) encode an identical peptide (S2/3) and the remaining isoforms encode unique mature proteins. Multiple sequence alignment of mature protein (123 amino acids) with unique substitutions highlighted in bold. The positions of the α-helices (red) and β-sheets (blue) are shown.

A)

	T1	T4	S1	T5	S3	S4	S2	S5
T1	100	92.68	94.31	93.5	94.31	94.31	92.68	93.5
T4	92.68	100	96.75	97.56	98.37	98.37	96.75	97.56
S1	94.31	96.75	100	98.37	98.37	98.37	98.37	97.56
T5	93.5	97.56	98.37	100	99.19	99.19	98.37	98.37
S3	94.31	98.37	98.37	99.19	100	100	98.37	99.19
S4	94.31	98.37	98.37	99.19	100	100	98.37	99.19
S2	92.68	96.75	98.37	98.37	98.37	98.37	100	99.19
S5	93.5	97.56	97.56	98.37	99.19	99.19	99.19	100

B)

Figure 3-5. Similarity of mature SSU peptide sequences. (A) Percent identity of mature SSU peptide sequences determined from a multiple sequence alignment (Fig. 3-4) using Clustal Omega (EMBL-EBI). **(B)** Phylogenetic tree produced from the multiple sequence alignment shows the evolutionary relationships between the mature peptide sequences.

3.2.2 Expression of rbcS isoforms in mature leaves

The relative expression of nine individual isoforms was measured by quantitative PCR (qPCR) to identify transcriptionally active members of the *rbcS* family. Primers were designed to amplify a *ca.* 200 bp region with the forward primer spanning the final exon-exon junction (Table 2-2). A full-length genomic sequence could not be amplified for *rbcS-T2,* but qPCR primers were designed based on the available partial-coding sequence (Gong et al., 2014). The expression of *L25 ribosomal protein* (*L25*) (GenBank: L18908) was used as an internal reference. *L25* is stably expressed during development and in tobacco plants exposed to abiotic stresses, including high temperature and low light (Schmidt and Delaney, 2010). Primer specificity was evaluated by PCR using a cDNA template and confirmed by a single dissociation peak in melt curve analysis (Fig. 3–6A, 6B). Standard curves were generated to quantify the amplification efficiency of each primer pair (Fig. 3–7). All of the primers

amplified the intended target with efficiencies ranging from 86–110%. The Pfaffl method was used to calculate gene expression relative to the internal reference gene and adjust for the variation in the primer efficiency (Pfaffl, 2001).

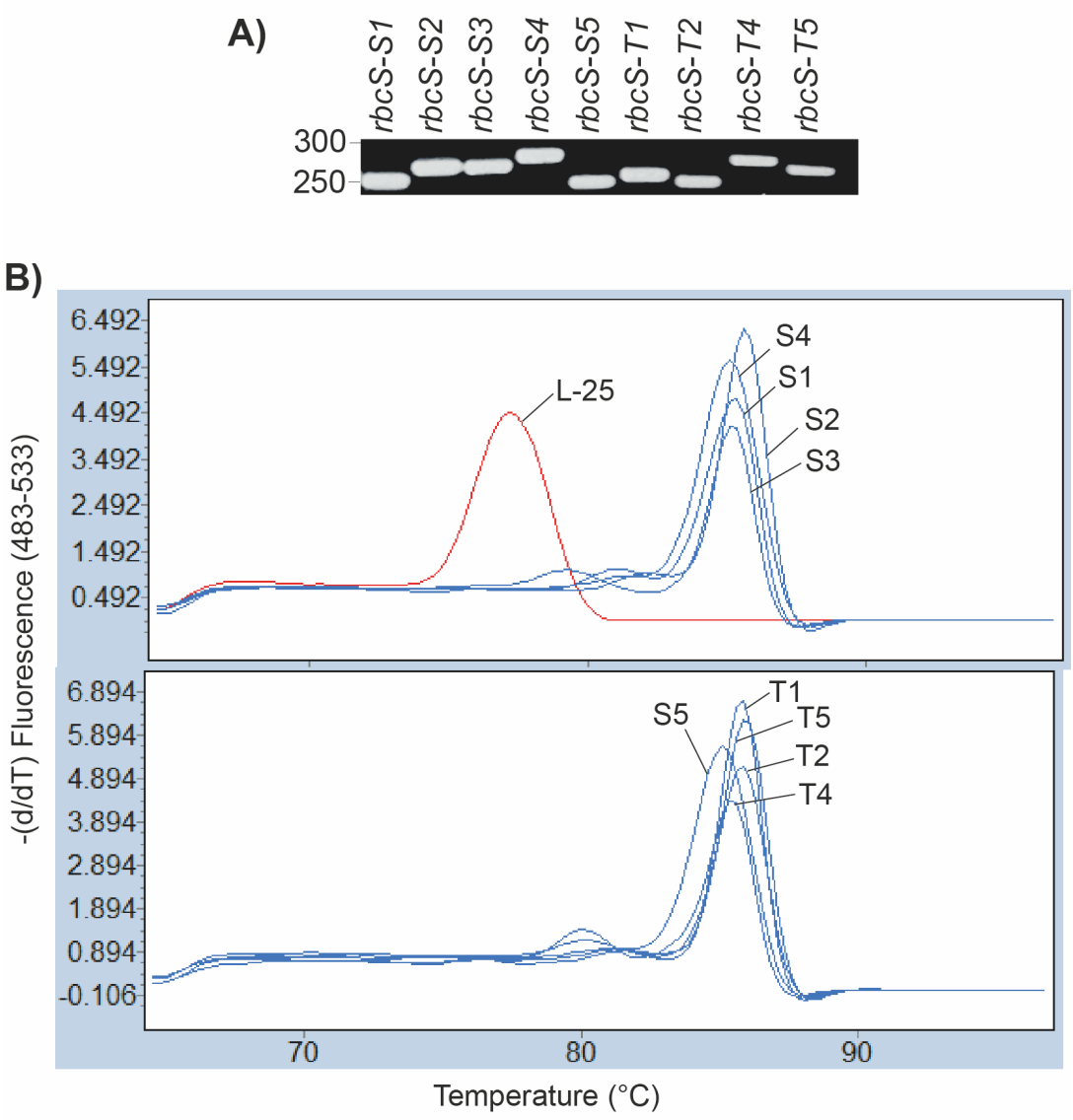


Figure 3-6. Specificity of qPCR primers (A) Analysis of qPCR primers by PCR of cDNA shows a single amplicon of the expected size for nine *rbcS* isoforms **(B)** Example of qPCR melt curve analysis from one biological replicate shows a single peak for nine *rbcS* isoforms (blue) and an internal reference gene (red).

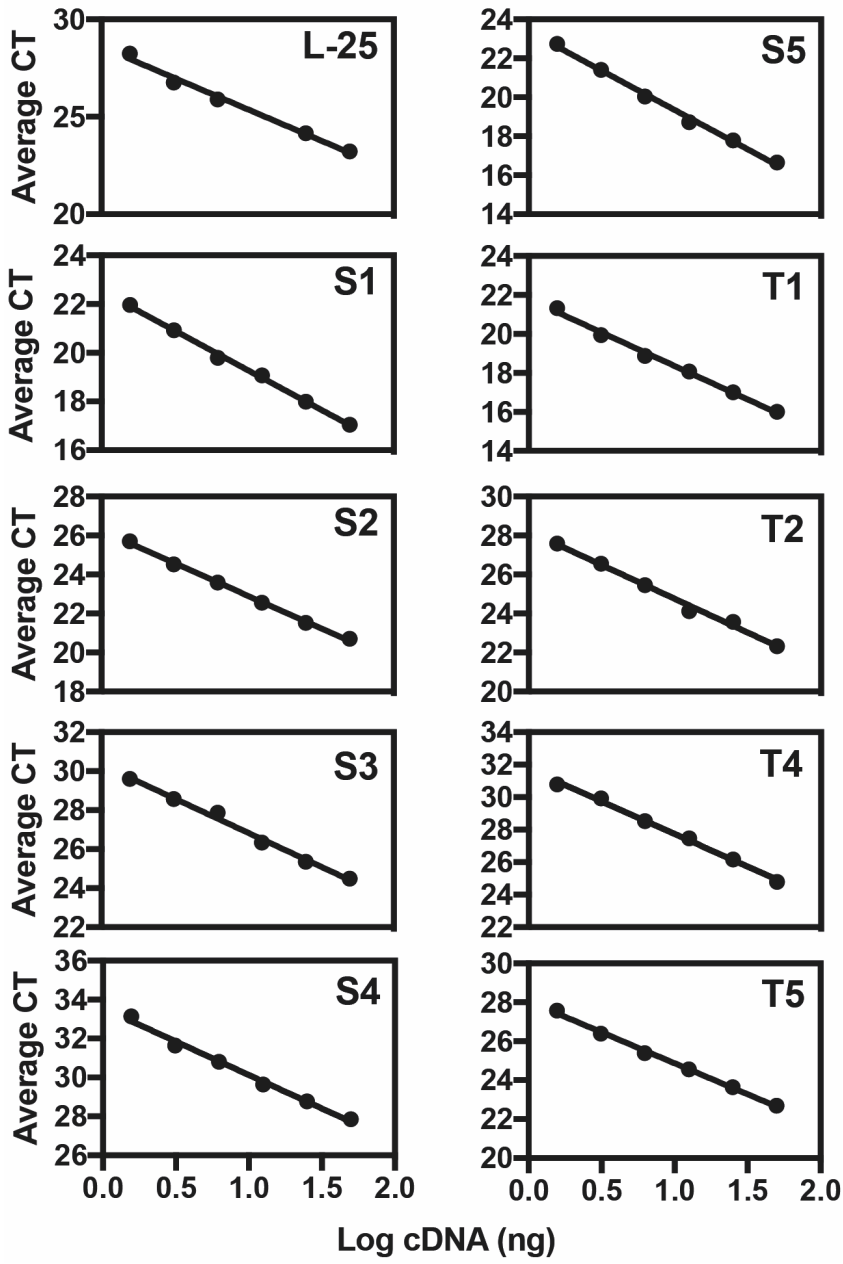


Figure 3-7. Standard curves to determine qPCR amplification efficiency. A serial dilution of cDNA was used to measure the efficiency of primers amplifying nine *rbcS* isoforms and an internal reference gene (L-25). Efficiency (E) was calculated from the gradient (m) as $E = 10^{(-1/m)}-1$ ($R^2 > 0.99$). Data show the average of three technical replicates.

cDNA samples from six-week-old tobacco plants were used to establish the relative expression of nine *rbcS* isoforms. Samples were harvested at a single time point in the middle of a 16-hour photoperiod. All of the primer pairs were included on a single plate to minimise variation between runs. The isoforms *rbcS-T1* and *rbcS-S1* accounted for *ca.* 60% of total *rbcS* transcripts (Fig. 3–8A). One other isoform, *rbcS-S5,* was abundantly expressed and accounted for 15% of total transcripts. The remaining isoforms were expressed at low levels (≤ 7%) in mature leaves. The maternally inherited S- isoforms accounted for 20% more of total transcripts than paternally inherited T- isoforms (Fig. 3–8B)

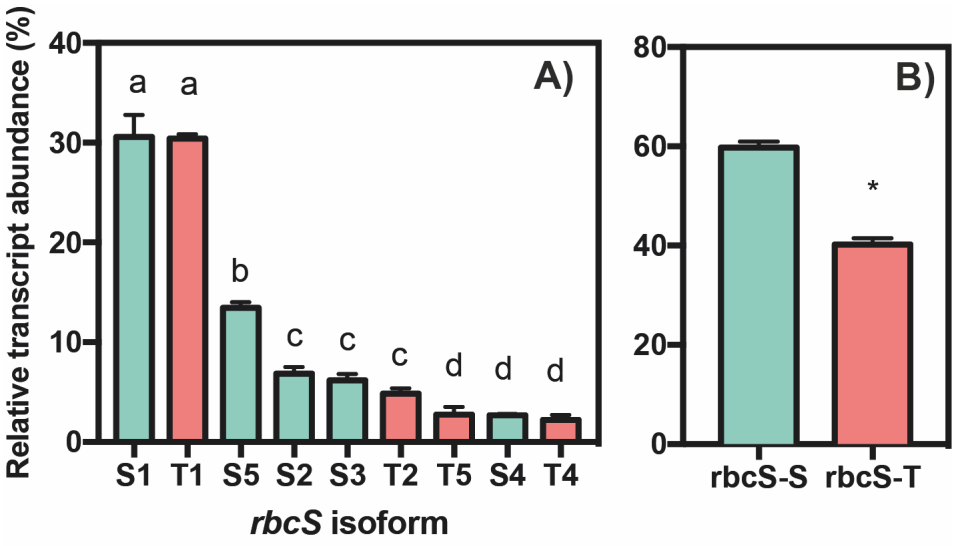


Figure 3-8. Relative abundance of mRNA transcripts encoding the tobacco *rbcS* family. **(A)** The abundance of individual *rbcS* transcripts was determined by qPCR using RNA isolated from the leaves of six-week-old tobacco plants. Bars indicate the mean abundance ±SEM (three biological replicates) as a percentage of total *rbcS* transcripts with letters indicating significant differences ($P < 0.05$) as determined by ANOVA followed by Tukey's HSD tests **(B)** The contribution of maternally inherited *rbcS-S* isoforms and paternally inherited *rbcS-T* isoforms to total transcripts. Bars indicate the same as described for (A) with significant differences indicated by an asterisk as determined by Student's t-test ($P < 0.0001$).

3.2.3 Expression of rbcS isoforms in light and dark-grown seedlings

The abundance of *rbcS* mRNA transcripts was measured in ten-day-old seedlings to identify the major isoforms expressed in response to light. Seedlings were grown under an irradiance of 150 µmol photons m^{-2} s^{-1} in a 12-hour photoperiod. Dark-grown (etiolated) seedlings were used to establish the baseline level of *rbcS* expression in the absence of light. The total amount of *rbcS* transcripts was significantly reduced in etiolated seedlings (Fig. 3–9). The major isoforms in light-grown seedlings, *rbcS-T1* and *rbcS-S1,* were also the most abundant in the dark. The accumulation of transcripts for most of the remaining isoforms was significantly lower, except for *rbcS-S4* that was equally abundant in light- and dark-grown seedlings.

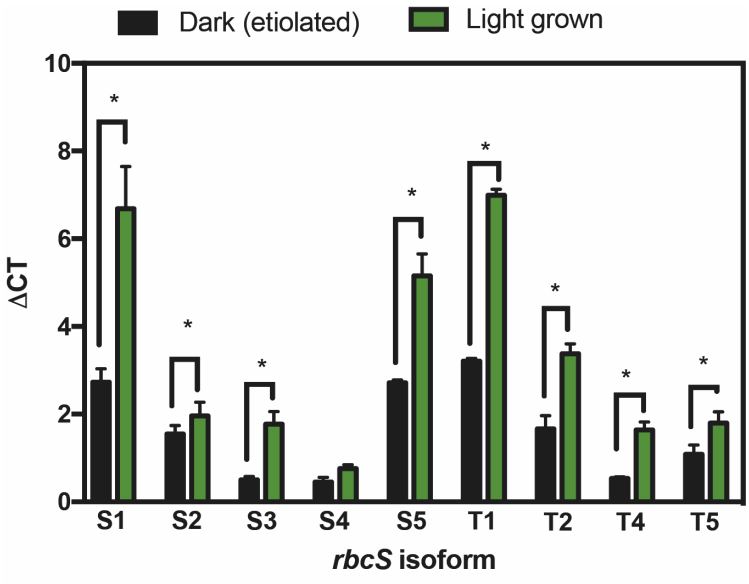


Figure 3-9. Relative abundance of mRNA transcripts encoding nine *rbcS* isoforms in tobacco. RNA was isolated from ten-day-old tobacco seedlings grown at an irradiance of 150 µmol photons m^{-2} s^{-1} (light-grown) or dark-grown (etiolated) seedlings. The relative abundance of transcripts (ΔCT) was calculated by normalising the CT values of each target to a constitutively expressed housekeeping gene. Light-grown seedlings were sampled in the middle of the photoperiod from four biological replicates, with twenty seedlings accounting for each replicate. Bars show the mean ±SEM (n = 4) with significant differences indicated by an asterisk ($P < 0.05$) as determined by *t*-test.

Oscillations in the accumulation of individual *rbcS* transcripts occurred in Arabidopsis plants that were grown in a diurnal cycle (Pilgrim and McClung, 1993; Stayton et al., 1989). Here, the diurnal rhythm of *rbcS* mRNA accumulation was established in ten-day-old tobacco seedlings. Seedlings grown at an irradiance of 150 μmol photons m^{-2} s^{-1} (12:12, L/D) were harvested two hours before dawn and periodically over eighteen hours. Three isoforms, *rbcS-S3*, *rbcS-S4* and *rbcS-T4*, were expressed at a consistent level (Fig. 3–10, 3-11). The abundance of mRNA transcripts encoding the remaining isoforms (*rbcS-S1, rbcS-S2, rbcS-S5, rbcS-T1, rbcS-T2* and *rbcS-T5*) peaked two hours before dawn. A slight decrease in the transcript level of *rbcS-S1*, *rbcS-S5, rbcS-T1* and *rbcS-T5* occurred within four hours of light exposure and was maintained throughout the photoperiod (Fig. 3–10, 3-11). A pre-dawn peak in the expression of six isoforms is consistent with circadian regulation of *rbcS* mRNA accumulation (Pilgrim and McClung, 1993). The absence of an oscillation in the accumulation of *rbcS-S3* and *rbcS-S4* and *rbcS-T4* transcripts suggests that these isoforms are regulated by additional environmental or developmental cues.

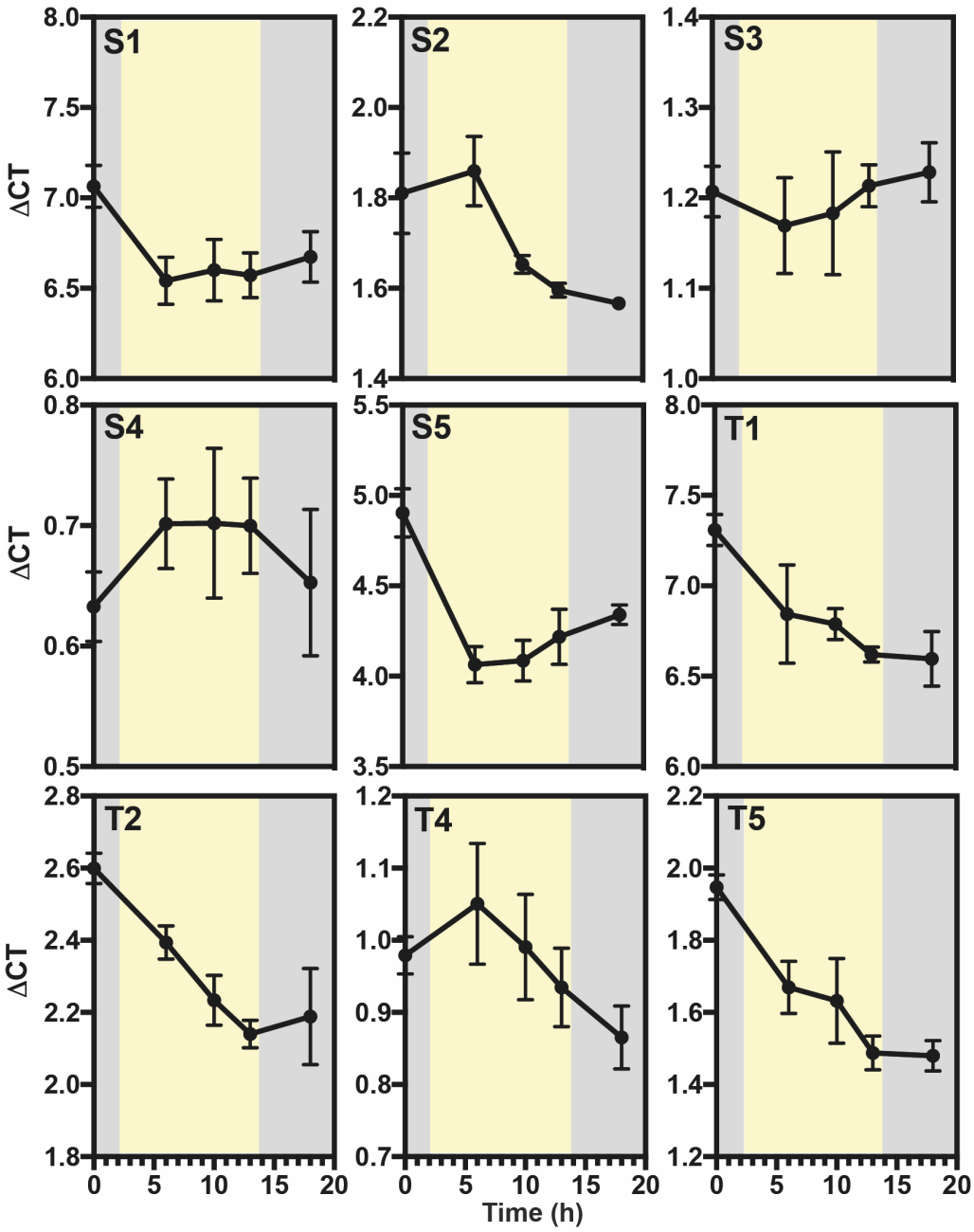


Figure 3-10. Time course of *rbcS* mRNA transcripts during a 12-hour photoperiod. RNA was isolated periodically from ten-day-old tobacco seedlings grown at an irradiance of 150 µmol photons m^{-2} s^{-1} in a 12:12 light/dark cycle. Dark time points are indicated by grey shading. The relative transcript level (ΔCT) was calculated as described in Figure 3–8. Each timepoint represents the mean ±SEM (n =6) from two independent experiments.

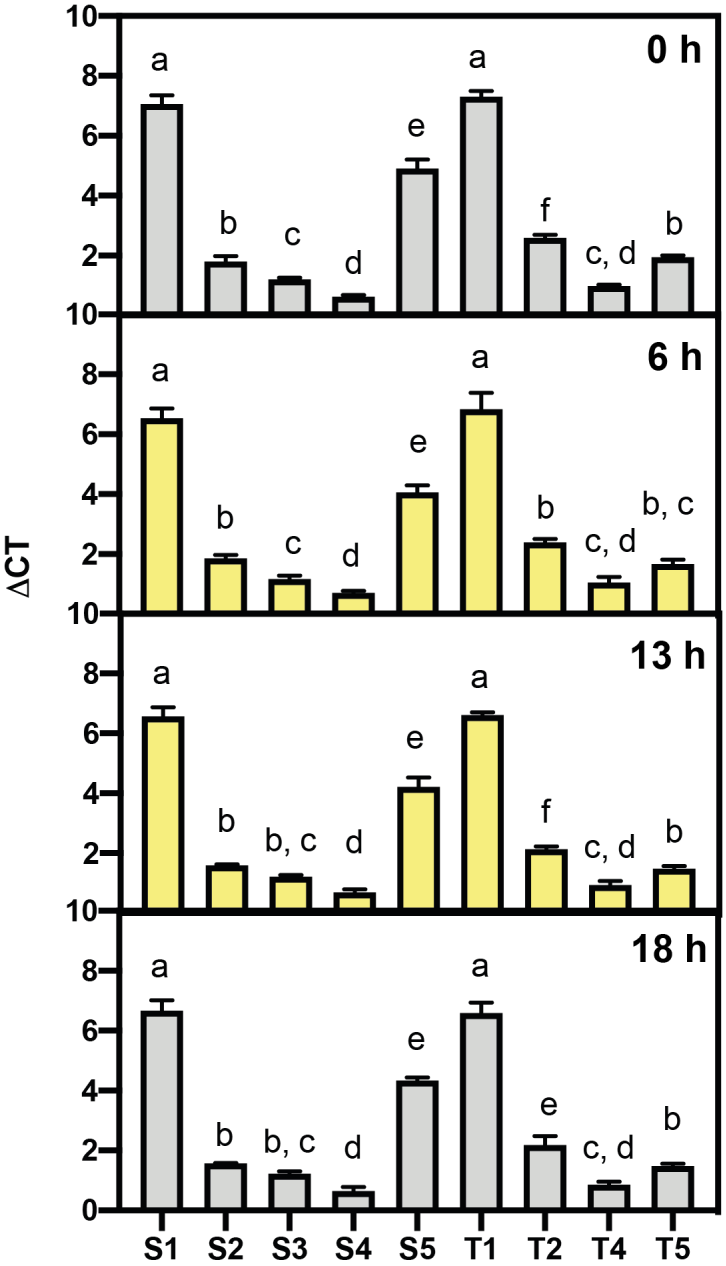


Figure 3-11. Time course of *rbcS* mRNA transcripts during a 12-hour photoperiod. RNA was isolated periodically from ten-day-old tobacco seedlings grown at an irradiance of 150 µmol photons m^{-2} s^{-1} in a 12:12 light/dark cycle. Dark time points are indicated by grey shading and the light period started after 2 hours and ended at 14 hours. The relative transcript level (ΔCT) was calculated as described in Figure 3–8. Each timepoint represents the mean ±SEM (n =6) from two independent experiments with different letters indicating significant differences (P = 0.05) between different isoforms as determined by ANOVA and Tukey's HSD test.

Dark-grown (etiolated) seedlings were exposed to twelve hours of light (150 μmol photons m^{-2} s^{-1}) to determine the pattern of light-induced mRNA accumulation in the absence of diurnal regulation. Transcripts for *rbcS-T1* and *rbcS-S5* increased during the photoperiod and remained relatively high after four hours of darkness (Fig. 3–12, 3–13). In the dark, the amounts of *rbcS-S1* and *rbcS-T1* transcripts were equivalent; however, after eleven hours of light exposure the amount of *rbcS-T1* transcript was significantly higher (Fig. 3–13). A similar pattern was observed for *rbcS-S1, rbcS-S2,* and *rbcS-S3* transcripts but a plateau was reached between four and eight hours of light exposure. Transcripts for *rbcS-T2* and *rbcS-T5* peaked after eleven hours of light and decreased more rapidly in the dark compared to the other isoforms. In particular, *rbcS-T2* transcripts after four hours of darkness were equivalent to the level measured before exposure to light (Fig. 3–13).The relative transcript level of two isoforms, *rbcS-S4* and *rbcS-T4,* did not increase during light exposure. As *rbcS-S4* and *rbcS-T4* are expressed in both dark- and light-grown seedlings additional environmental or developmental cues could regulate the expression of these isoforms.

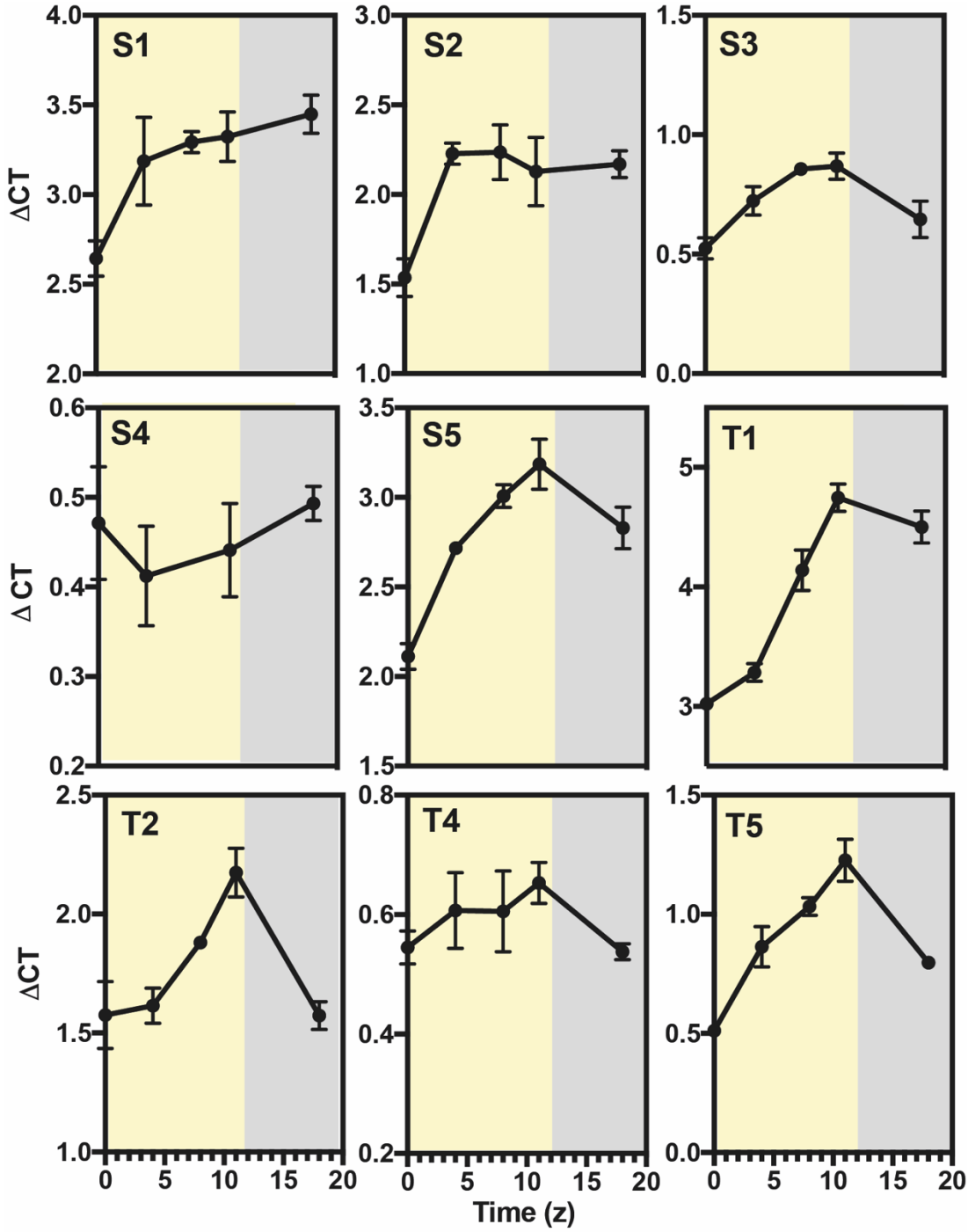


Figure 3-12. Accumulation of *rbcS* mRNA transcripts in response to light. RNA was isolated from ten-day-old dark grown (etiolated) seedlings exposed to an irradiance of 150 µmol photons m^{-2} s^{-1} for 12 hours. Dark time-points are indicated by grey shading. The relative transcript level (ΔCT) was calculated as described in Figure 3–8. Each timepoint represents the mean ±SEM (n =6) from two independent experiments.

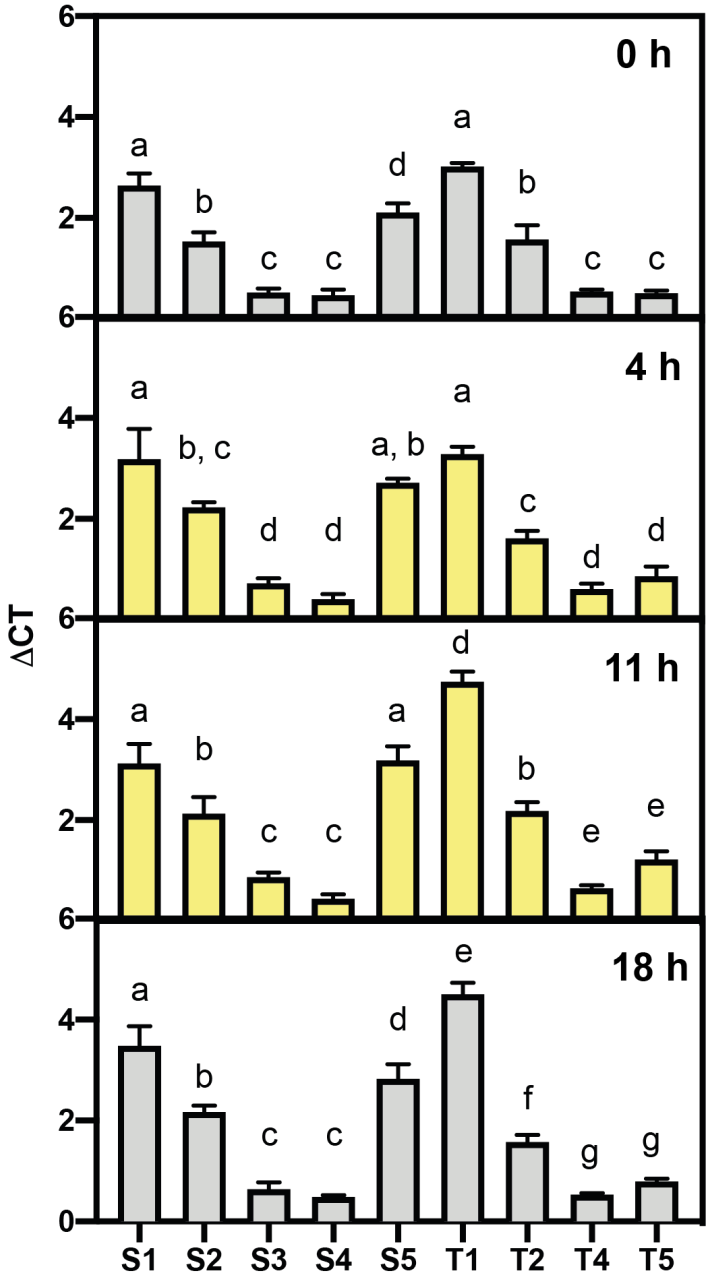


Figure 3-13. Accumulation of *rbcS* mRNA transcripts in response to light. RNA was isolated from ten-day-old dark grown (etiolated) seedlings exposed to an irradiance of 150 µmol photons m^{-2} s^{-1} for 12 hours. Dark time points are indicated by grey shading and the light period started after 2 hours and ended at 14 hours. The relative transcript level (ΔCT) was calculated as described in Figure 3–8. Each timepoint represents the mean ±SEM (n =6) from two independent experiments with different letters indicating significant differences ($P = 0.05$) between different isoforms as determined by ANOVA and Tukey's HSD test.

Individual *rbcS* isoforms are differentially regulated in response to RL, BL and FRL (Dedonder et al., 1993; Wehmeyer et al., 1990). Classical phytochrome responses can be characterised by FRL reversibility of RL-induced transcription. Etiolated seedlings were exposed to four different light treatments to establish the role of RL- and BL-mediated regulation of *rbcS* mRNA accumulation in tobacco. Ten-day old seedlings were exposed to a single pulse (90 μmol photons m^{-2} s^{-1} for 10 minutes) of RL (630-670 nm), FRL (720-740 nm) or BL (440-460 nm). An additional group (RL-FRL) was exposed to FRL pulse following the RL treatment. Seedlings were harvested every two hours after the light treatment over eight hours.

The accumulation of mRNA transcripts for all nine *rbcS* isoforms increased in response to RL (Fig. 3–14A, 15A, 16A). Transcripts for *rbcS-S1* peaked four hours after the light pulse (Fig. 3–14A) and the fold change was significantly higher than for the other eight isoforms (Fig. 3–17). A similar pattern of expression was found for *rbcS-T1*, which increased by 1.2-fold. The accumulation of *rbcS-S5* transcripts increased by a total of 1-fold compared to the baseline in the dark and was significantly lower than *rbcS-T1* and *rbcS-S1* (Fig. 3–17). Although *rbcS-S4* and *rbcS-T4* transcripts did not accumulate in response to continuous white light (Fig. 3–13), the RL pulse induced a significant fold change in both isoforms (Fig. 3–15A, 17A). Transcripts for *rbcS-S3*, *rbcS-T2* and *rbcS-T4* increased by *ca.* 0.5-fold reaching a statistical maximum between six- and eight- hours. After the FRL treatment, there was less transcript accumulation for *rbcS-S1*, *rbcS-T1*, *rbcS-S5,* and *rbcS-T5* than after the RL treatment (Fig. 3–18), and all of the isoforms reached maximum expression within two hours (Fig. 3–14B, 16B, 17B). The extent of FRL reversibility in response to RL (RL-FRL) was highly variable between the three predominant isoforms. The abundance of *rbcS-S1* transcripts 8 hours after the RL-FRL treatment was significantly lower than RL alone but higher than after the FRL treatment (Fig. 3–18). Transcripts for *rbcS-T1* were more strongly supressed by the RL-FRL treatment and

not significantly different than after FRL exposure (Fig. 3–18). In contrast, the accumulation *rbcS-S5* mRNA following RL-FRL was equivalent to the RL response (Fig. 3–18). Similarly, transcripts *rbcS-T5* were equivalent between the RL and RL-FRL conditions eight hours after exposure (Fig. 3–18).

The accumulation of *rbcS* mRNA transcripts was observed following a BL pulse for six isoforms (*rbcS-S1*, *rbcS-T1*, *rbcS-S5, rbcS-S3* and *rbcS-T4*). Of the six, the expression of *rbcS-S1* and *rbcS-T1* increased the most in response to BL (Fig. 3–14D, 3–17). The fold change of *rbcS-T1* was equivalent between the RL and BL treatments, while *rbcS-S1* transcripts increased comparatively less after the BL pulse (Fig. 3–18). The accumulation of *rbcS-S5* and *rbcS-T2* transcripts after BL exposure was also significantly lower than after the RL treatment (Fig. 3–16D, Fig. 3–18). Transcripts for the remaining two isoforms (*rbcS-S3* and *rbcS-T4*) were not significantly different between the BL and RL treatments (Fig. 3–16D, Fig. 3–18).

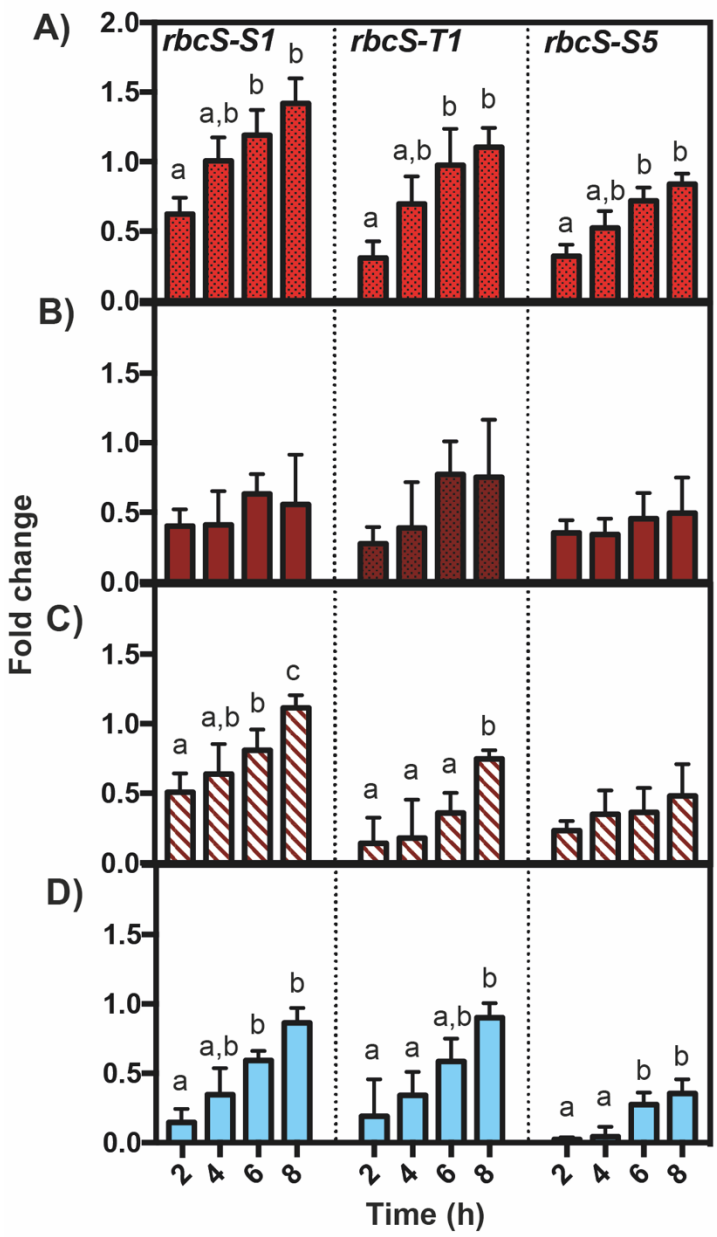


Figure 3-14. Fold change of Rubisco small subunit transcripts (*rbcS-S1*, *rbcS-T1*, *rbcS-S5*) in response to different wavelengths of light. The abundance of *rbcS-S1*, *rbcS-T1* and *rbcS-S5* transcripts in ten-day-old dark-grown (etiolated) tobacco seedlings exposed to a pulse (90µmol m^{-2} s^{-1} for 10 minutes) of red (A), far-red (B), red followed by far-red (C) or blue (D) light. The x-axis shows time as hours after exposure to the light pulse. Fold change was calculated relative to etiolated seedlings in the dark. Bars represent the mean ±SEM (n=4-6). from two independent experiments with different letters indicating significant differences ($P<0.05$) determined by ANOVA and Tukey's HSD test.

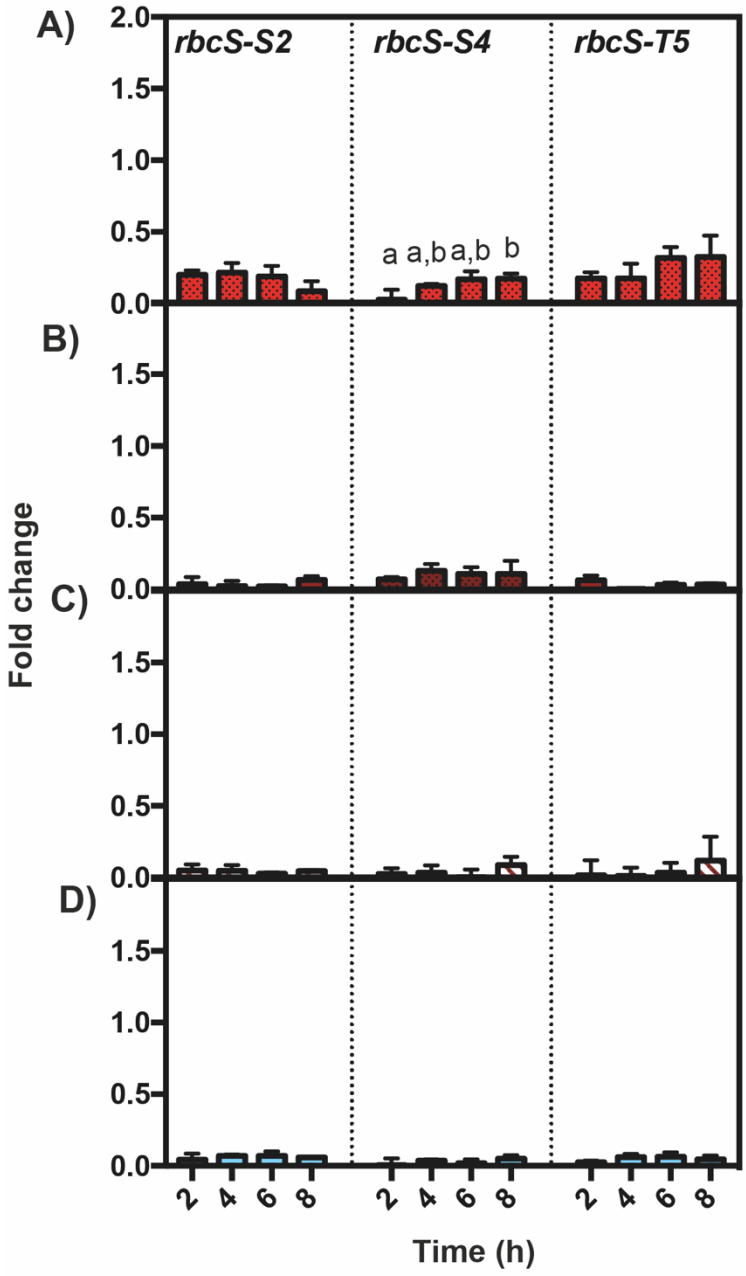


Figure 3-15. Fold change of Rubisco small subunit transcripts (*rbcS-S2, rbcS-S4, rbcS-T5*) in response to different wavelengths of light. The abundance of *rbcS-S2, rbcS-S4* and *rbcS-T5* transcripts in ten-day-old dark-grown (etiolated) tobacco seedlings exposed to a pulse (90µmol m^{-2} s^{-1} for 10 minutes) of red (A), far-red (B), red followed by far-red (C) or blue (D) light. The x-axis shows time as hours after exposure to the light pulse. Fold change was calculated relative to etiolated seedlings in the dark. Bars represent the mean ±SEM (n=4-6). from two independent experiments with different letters indicating significant differences ($P<0.05$) determined by ANOVA and Tukey's HSD test.

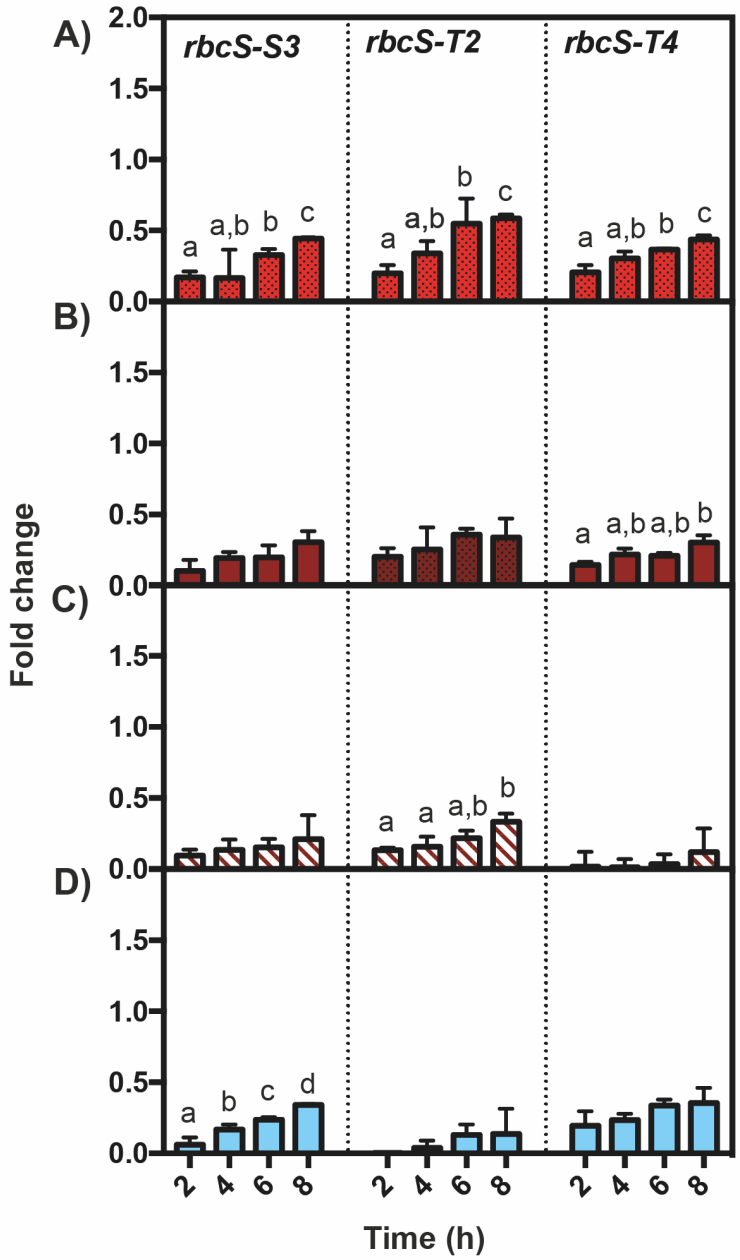


Figure 3-16. Fold change of Rubisco small subunit transcripts (*rbcS-S3*, *rbcS-T2*, *rbcS-T4*) in response to different wavelengths of light. The abundance of *rbcS-S3, rbcS-T2, rbcS-T4*) transcripts in ten-day-old dark-grown (etiolated) tobacco seedlings exposed to a pulse (90µmol m^{-2} s^{-1} for 10 minutes) of red (A), far-red (B), red followed by far-red (C) or blue (D) light. The x-axis shows time as hours after exposure to the light pulse. Fold change was calculated relative to etiolated seedlings in the dark. Bars represent the mean ±SEM (n=4-6). from two independent experiments with different letters indicating significant differences ($P<0.05$) determined by ANOVA and Tukey's HSD test.

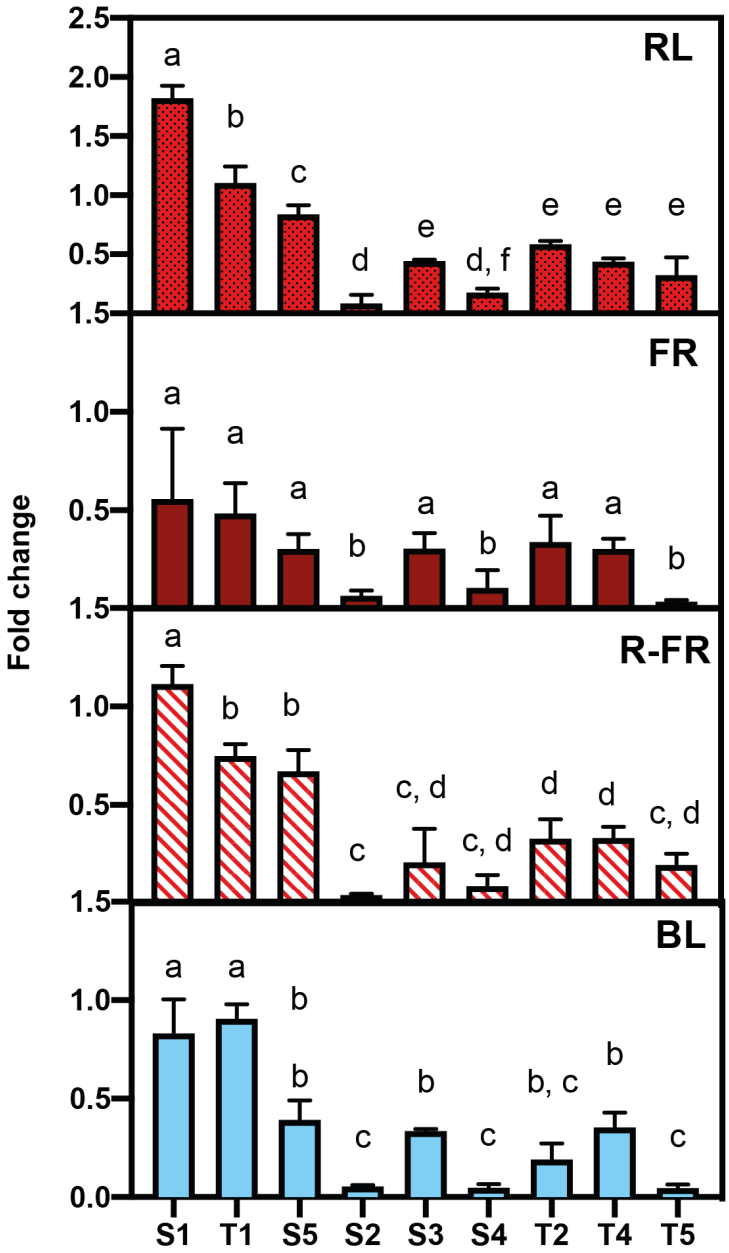


Figure 3-17. Fold change of Rubisco small subunit transcripts 8 hours after exposure to different wavelengths of light. The abundance of *rbcS* transcripts in ten-day-old dark-grown (etiolated) tobacco seedlings exposed to a pulse (90µmol m^{-2} s^{-1} for 10 minutes) of red (RL), far-red (FR), red followed by far-red (R-FR) or blue (BL) light. Fold change was calculated relative to etiolated seedlings in the dark. Bars represent the mean ±SEM (n=4-6). from two independent experiments with different letters indicating a significant difference ($P<0.05$) between different *rbcS* isoforms in each light condition as determined by ANOVA and Tukey's HSD test.

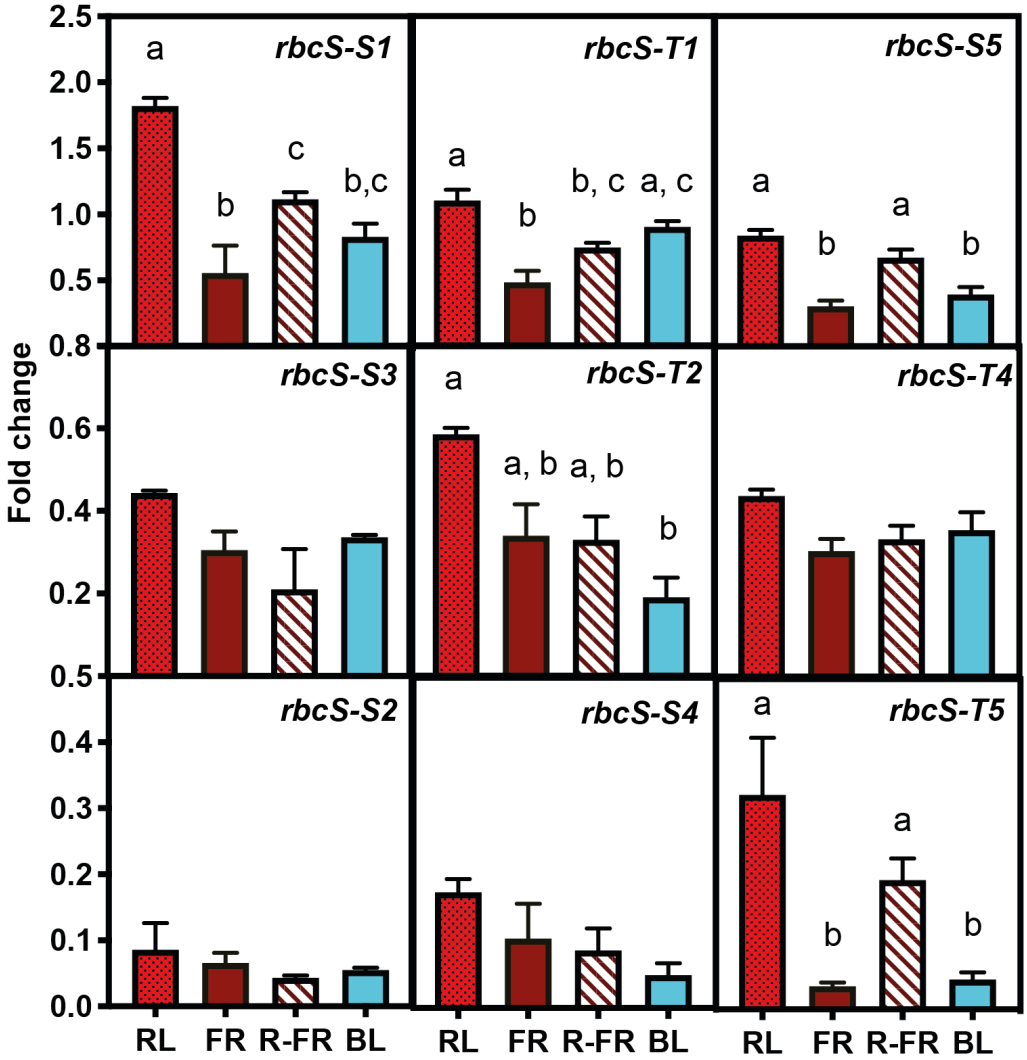


Figure 3-18. Fold change of Rubisco small subunit transcripts 8 hours after exposure to different wavelengths of light. The abundance of *rbcS* transcripts in ten-day-old dark-grown (etiolated) tobacco seedlings exposed to a pulse (90µmol m^{-2} s^{-1} for 10 minutes) of red (RL), far-red (FR), red followed by far-red (R-FR) or blue (BL) light. Fold change was calculated relative to etiolated seedlings in the dark. Bars represent the mean ±SEM (n=4-6). from two independent experiments with different letters indicating a significant difference ($P<0.05$) among the different light conditions for each isoform as determined by ANOVA and Tukey's HSD test.

3.3 Discussion

Thirteen partial sequences of tobacco *rbcS* genes were previously described that had different gene structures and nucleotide coding sequences (Gong et al., 2014). Allopolyploidy results in the inheritance of two nuclear genomes, while a single chloroplast genome is inherited from the maternal progenitor. Duplicated genes can undergo sequence or regulatory changes that include gene conversion, concerted evolution, and gene loss or silencing of expression (Wendel, 2000). Interactions between nuclear- and chloroplast- encoded gene products can influence the evolution and regulation of duplicated gene products. For example, the *Gossypium rbcS* family has extensive paternal to maternal gene conversions that have occurred post-polyploidization (Gong et al., 2012). Gene conversion of tobacco *rbcS* isoforms is low in comparison to more ancient polyploids like *Gossypium* (Gong et al., 2014). However, more *rbcS* genes are present at the polyploid level than in the diploid progenitors.

Three pairs of paralogous *rbcS* genes have undergone gene conversion in tobacco: *rbcS1a*/*rbcS1b, rbcST3a*/*rbcST3b,* and *rbcST4a*/*rbcSTb* (Gong et al., 2014). Each pair of isoforms share a high percent similarity (97 to ≥ 99 %) that includes the usually divergent intron sequences. Unique regions of the genome could not be identified for *a* and *b* homologs because of high sequence similarity. Up to 90% of the tobacco genome has been assembled by whole genome shotgun (WGS) *de novo* sequence assembly (Sierro et al., 2014; Edwards et al., 2017). Genomic complexity can be lost during the assembly of WGS sequences that contain identical repeats resulting in the loss of duplicated genes (Alkan et al., 2011). The most recent tobacco genome assembly retained a higher number of genes encoded by families (Edwards et al., 2017). However, the sequences surrounding *rbcS-S1*, *rbcS-T3* and *rbcS-T4* could not be resolved (denoted by a string of 'N's) suggesting the presence of an identical sequence during assembly. Although it was not possible to discriminate between the genomic sequences of the *rbcS-S1*, *rbcS-T3,* and *rbcS-T4 a* and *b* homologs in this study, the high

coding sequence similarity is expected to allow the combined expression of the isoform pairs to be measured by qPCR. Improved assembly of repetitive sequences and duplicated regions of the tobacco genome will enable individual characterisation of the *a* and *b* homologs.

Two *rbcS* isoforms, *rbcS-S1* and *rbcS-T1*, have a unique gene structure due to the presence of a third intron. The remaining isoforms have two introns that occur in a common position for most isoforms in dicot species. The third intron in *rbcS-S1* and *rbcS-T1* occurs at a highly conserved region of the mature protein between amino acids 65 and 66. A third intron is also present at the same position, but only in a single isoform, in other *Solanaceae* species including potato (Fritz et al., 1993; Wolter et al., 1988), tomato (Sugita et al., 1987), and petunia (Dean et al., 1987). The three-intron isoforms are unlinked to the rest of the *rbcS* family within individual species. Notably, three-intron isoforms account for the majority of *rbcS* transcripts in the mature leaves of tobacco (*rbcS-S1* and *rbcS-T1*), potato (St*rbcS1*), and petunia (SSU301) (Dean et al., 1985; Fritz et al., 1993). The duplication of the *rbcS* three-intron sub-family is at least as old as the *Solanaceae* family (Dean et al., 1989; Weeks et al., 2007). Phylogenetic analysis of Sl*rbcS2*, SSU301, St*rbcS1* and *rbcS-S1* identified common regulatory motifs in the 5' UTR of each gene (Weeks et al., 2007). The non-coding regions of the two orthologous genes in tomato and potato share an unusually high (>83%) nucleotide identity (Fritz et al., 1993). Based on the divergence of the three-intron gene orthologs in other *Solanaceae* species from within the respective gene families, it is unsurprising that *rbcS-S1* and *rbcS-T1* have the most intraspecific variation and are the most highly expressed isoforms.

The two most divergent isoforms, *rbcS-S1* and *rbcS-T1,* account for *ca.* 60% of total *rbcS* transcripts in tobacco leaves. A posthoc analysis of RNA-seq data generated from tobacco leaves found that *rbcS-T1* and *rbcS1* (rbcS1*a* and *rbcS1b* isoforms combined) accounted for ≥ 80% of the relative abundance of transcripts (Lin et al., 2019). Although the coding sequences

of *rbcS1a* and *rbcS1b* differ by only two nucleotides, it is possible that only one isoform was amplified by the qPCR primers in this study. Therefore, the contribution of *rbcS-S1* isoforms could be underestimated. The contribution of *rbcS-S5* transcripts was also significantly lower than reported in this study (5% compared to 15%). Genetic variation between cultivars of tobacco could account for differences in the accumulation of mRNA transcripts. Furthermore, different developmental stages and growth conditions could influence the expression of individual *rbcS* isoforms. We measured the relative expression of *rbcS* isoforms in the leaves of six-week-old plants grown in natural greenhouse conditions. In contrast, Sierro et al. (2014) isolated RNA from sixteen-week old plants grown axenically on MS medium. However, our results are consistent with the previous suggestion that there is a biased expression of maternally inherited *rbcS* isoforms in polyploid species (Gong et al., 2014, 2012). It is possible that this bias could occur because of preferential expression of nuclear- and the maternally inherited chloroplast products (i.e. *rbcL*). In the current study, the *rbcS-S* isoforms accounted for 60% of total transcripts in tobacco leaves. However, potential overestimation of *rbcS-S5* and the lack of data for *rbcS-T3* must be taken into consideration. Nevertheless, transcripts for the *rbcS-S* isoforms are still considerably more abundant, even taking a conservative estimate for the *rbcS-S5* level into account.

Diurnal oscillations in the mRNA accumulation of six out of nine *rbcS* isoforms (*rbcS-S1, rbcS-S2, rbcS-S5, rbcS-T1, rbcS-T2,* and *rbcS-T5*) occurred in light-grown seedlings. The highest peak in gene expression of all the isoforms was measured close to dawn and decreased over the photoperiod. Similarly, mRNA transcripts encoding the *rbcS* family in Arabidopsis and pea (*Pisum sativum*) accumulated before the start of the light period (Kloppstech, 1985; Piechulla et al., 1986; Piechulla and Gruissem, 1987; Pilgrim and McClung, 1993). The pattern of transcript accumulation persisted when Arabidopsis plants were transferred to continuous dark or light conditions, which suggests circadian regulation (Pilgrim and McClung, 1993). In contrast, diurnal oscillations in total *rbcS* transcript accumulation were not identified in

tobacco (Paulsen and Bogorad, 1988). Here, the magnitude of diurnal oscillations differed between individual *rbcS* isoforms in tobacco. Transcripts for three isoforms, *rbcS-S1*, *rbcS-T1* and *rbcS-S* decreased slightly during the photoperiod while a clearer oscillation was detected for the minor isoforms (*rbcS-S2*, *rbcS-T2* and *rbcS-T5*). It would be necessary to measure expression over a longer period (e.g., 24 hours) to confirm a diurnal pattern of expression for the major isoforms. Furthermore, transferring plants to continuous dark or light conditions would provide a better understanding of the diurnal and circadian regulation of individual *rbcS* isoforms in tobacco.

The analysis of steady-state mRNA levels in etiolated seedlings demonstrates that individual *rbcS* isoforms are regulated by distinct photo- and developmental responses. All nine isoforms measured here were not completely absent but expressed at a basal level in the dark, indicating that *rbcS* expression is not solely light-regulated. The magnitude of *rbcS* expression in dark-grown seedlings varies between species and individual *rbcS* isoforms. In etiolated seedlings of maize (Sheen and Bogorad, 1986), *Amaranth* (Berry et al., 1985), and potato (Fritz et al., 1991) *rbcS* mRNA was detectable in the dark. However, in *Lemna* (Silverthorne et al., 1990) and soybean (Berry-Lowe and Meagher, 1985) there was little accumulation of *rbcS* mRNA compared to light-grown plants. Individual *rbcS* isoforms were differentially expressed in light- and dark-grown Arabidopsis (Dedonder et al., 1993) and tomato seedlings (Sugita and Gruissem, 1987). One Arabidopsis *rbcS* isoform, *rbcS2B,* was equally abundant in five day old light- and dark-exposed etiolated seedlings (Dedonder et al., 1993). However, light modulated the expression of *rbcS2B* in ten- to fourteen-day-old seedlings. Similarly, in the current study the relative transcript level of the tobacco isoform *rbcS-S4* was not significantly different between light-grown seedlings and dark- or light-exposed etiolated seedlings. Therefore, developmental programmes likely influence the expression of *rbcS-S4* in the dark. Other isoforms appeared to be regulated by a combination of light- and developmental cues. For example, the transcripts of two isoforms (*rbcS-T4* and *rbcS-T5*) accumulate less in

etiolated seedlings than in light-grown seedlings. However, continuous white light did not increase the abundance of *rbcS-T4* and *rbcS-S4* transcripts in etiolated seedlings. This suggests that light modulates the developmental cues that regulate *rbcS-T4* and *rbcS-S4* expression in light-grown seedlings. The remaining isoforms are strongly light-regulated in etiolated tissues but differ in the rate of steady-state mRNA accumulation. Three isoforms (*rbcS-S1, rbcS-S5* and *rbcS-S2*) reach steady-state mRNA levels within four hours of continuous light, while *rbcS-T1, rbcS-S3* and *rbcS-T2* have slower rates of accumulation (between four and eleven hours). Increased accumulation of total *rbcS* mRNA is mostly controlled by the rate of transcription in tobacco (Wehmeyer et al., 1990). However, differences in mRNA degradation can affect the steady-state mRNA level of individual *rbcS* isoforms in potato (Fritz et al., 1991). Transcripts for individual tobacco *rbcS* isoforms also decrease at different rates in the dark. The transcript levels of the isoforms with the highest fold change, *rbcS-S1, rbcS-T1* and *rbcS-S5*, do not decrease significantly after four hours of darkness. In contrast, mRNAs encoding *rbcS-S2, rbcS-T2* and *rbcS-T5* decrease rapidly after light exposure.

Classical phytochrome responses can be measured by the photo-reversibility of RL-induced gene expression by an FRL pulse. The photo-reversibility of different phytochrome (phy) family members vary in response to FRL. For example, in Arabidopsis phyB is rapidly activated by RL and reversed by FRL, while phyA is maximal under continuous FRL (Smith, 2000). Total *rbcS* mRNA accumulates in etiolated tobacco seedlings after a five-minute pulse of RL or FRL (Wehmeyer et al., 1990). An FRL pulse after the RL treatment partially reverses the accumulation of RL transcripts, but not to the level induced by FRL alone. Measuring the steady-state mRNA levels of individual *rbcS* isoforms in seedlings exposed to the same light treatments demonstrates that the tobacco *rbcS* gene family is differentially regulated by phytochrome. Transcripts for six out of nine isoforms (excluding *rbcS-S2, rbcS-S4* and *rbcS-T5*) accumulated in response to the low levels of P_{FR} induced by an FRL pulse. The steady-state mRNA levels for all nine isoforms increased following an RL pulse. However, the extent

of suppression following a subsequent FRL pulse differed between isoforms. Transcripts for *rbcS-S1* and *rbcS-T1* were suppressed to the same or a slightly higher level as induced by FRL alone. However, the accumulation of RL-induced *rbcS*-S5 and *rbcS-T5* transcripts was not significantly supressed by the FRL pulse. The degree of phytochrome regulation varies between species and within *rbcS* families. Phytochromes differentially regulate individual *rbcS* family members in Arabidopsis (Dedonder et al., 1993) and pea (Fluhr and Chua, 1986). One Arabidopsis isoform, *rbcS1B* is regulated in a phytochrome-independent manner. However, distinct phytochrome-responses of individual isoforms vary in pea depending on the developmental stage (Fluhr and Chua, 1986). Two *rbcS* isoforms are sensitive to RL in etiolated seedlings, but not in mature leaves, which also require BL (Fluhr and Chua, 1986). The requirement for both BL and RL suggests that photoreceptors other than phytochrome regulate the expression of two pea isoforms in combination with developmental cues.

Transcripts for most of the remaining isoforms accumulate less in BL than in RL. Transcript accumulation in response to BL may be a result of stimulating low levels of P_{FR} (Schäfer and Haupt, 1983). However, *rbcS-T1* transcripts are equally abundant after exposure to BL or RL. Similarly, the accumulation of two *rbcS* transcripts in pea is equivalent between BL and RL (Fluhr and Chua, 1986). It is unclear whether the accumulation of *rbcS* transcripts in response to BL is phytochrome-independent. However, the accumulation of mRNA transcripts one Arabidopsis isoform, *rbcS-3B,* increases in response to BL in seedlings with phytochrome saturated by RL (Dedonder et al., 1993). It is likely that *rbcS-3B* is regulated by a BL receptor as well as phytochrome. Other isoforms require a combination of signals for expression. For example, transcripts for the Arabidopsis *rbcS1B* accumulate in response to white light, but not RL or BL alone. Similarly, BL significantly increases the expression of total *rbcS* transcripts compared to RL in bean (*Phaseolus vulgaris*) (Sawbridge et al., 1993). Further experiments are needed to understand if the BL responses of the *rbcS* isoforms, particularly *rbcS-T1*, are

mediated by phytochrome or a BL receptor. However, this work provides a preliminary understanding of the differential expression of individual *rbcS* isoforms in response to BL.

This work has provided the first characterisation of the majority of *rbcS* family in tobacco. Improved coverage of the tobacco genome assembly will assist future efforts to distinguish between the *a* and *b* homologs of three isoforms (*rbcS-S1, rbcS-T3*, *rbcS-T4*). Furthermore, identifying the locus expressing trichome Rubisco (*rbcS-t*) will enable comparative studies with the *rbcS* isoforms reported in this work (Laterre et al., 2017). Characterising the response of individual *rbcS* isoforms to light and spectral quality is important for understanding the regulation of Rubisco content in different environments. Furthermore, identifying the major *rbcS* isoforms that contribute to tobacco Rubisco is useful for future engineering efforts that will require the replacement of endogenous SSUs.

Chapter 4

Engineering the tobacco *rbcS* multigene family with the RNA-guided endonuclease SpCas9

4.1 Introduction

Modifying Rubisco in the leaves of *Nicotiana tabacum* (tobacco) is a useful approach for understanding structural requirements of Rubisco and the relationship between Rubisco content, photosynthesis and growth. Attempts to engineer hybrid Rubisco enzymes *in planta* have been useful in developing our understanding of Rubisco assembly, the impact of different subunits on the catalytic characteristics of Rubisco and the subsequent effects on photosynthesis and growth (section 1.4.1). In tobacco, the role of the SSU has been examined only in terms of its impact Rubisco content. Targeting *rbcS* isoforms with antisense mRNA provided an insight into the extent of Rubisco limitation on photosynthetic rate and plant growth (Stitt and Schulze, 1994). Unlike the LSU, replacing the native SSU family with heterologous SSU(s) (e.g., from Rubisco enzymes that have different catalytic characteristics) has not yet been achieved. In plants it remains a grand challenge to remove multiple native SSUs (e.g., by T-DNA insertion or mutagenesis), particularly in polypoid species that have multiple *rbcS* homeologs. Tobacco has at least thirteen *rbcS* isoforms (Gong et al., 2014), and so far, it has not been possible to disrupt the expression of individual *rbcS* isoforms. RGENs including *SpCas9* (Jinek et al., 2012) have emerged as tools that enable the editing of multiple genes in polyploids with relative ease (Morineau et al., 2017; Wang et al., 2018, 2014). Therefore, RGENs present a unique opportunity to knock-out multiple members of the tobacco *rbcS* family.

Previously, generating plants with a reduced Rubisco content has provided a means to evaluate the extent of Rubisco limitation on CO_2 assimilation rates and growth (Furbank and Taylor, 1995; Stitt and Schulze, 1994). One of the first studies to suppress the expression of *rbcS* genes, and thus decrease Rubisco content, employed an antisense approach in tobacco plants (Rodermel et al., 1988). Subsequently, Hudson et al. (1992) applied the same technique to generate plants with a stronger suppression of *rbcS*. Rubisco could be decreased to 40% of wild-type levels before photosynthesis and growth were significantly impaired in transgenic plants (Masle et al., 1993; Quick et al., 1991c, 1991a; Stitt et al., 1991). Further studies with the antisense lines provided an understanding of the impact of Rubisco limitation in response to nitrogen availability (Fichtner et al., 1993), variable light intensities (Lauerer et al., 1993), and high temperature (Krapp et al., 1994). Photosynthesis and growth were restored to wild-type levels by complementing an antisense-SSU tobacco mutant with a chloroplast-encoded *rbcS* from tobacco (Dhingra et al., 2004). Similarly, antisense rice (*Oryza sativa L.*) plants (Makino et al., 1997) were complemented with a divergent *rbcS* isoform that altered the catalytic properties of Rubisco (Morita et al., 2014). Successful complementation of antisense lines highlights the value of genotypes with reduced Rubisco content as a model to express non-native SSUs.

Although antisense approaches have advanced our understanding of the extent of Rubisco limitation on photosynthesis, this technique has several drawbacks. The effectiveness of *rbcS* suppression varies between plants, tissues, and developmental stages (Quick et al., 1991c). Therefore, it is necessary to determine the Rubisco content of each plant to analyse growth and photosynthesis. The extent of suppression in the progeny can also vary, with a significant loss of suppression reported in the T_1 generation of antisense *rbcS* wheat (Mitchell et al., 2004). An Arabidopsis T-DNA insertion line (*1a3b*) has *ca.* 30% of wild-type Rubisco (Izumi et al., 2012) and is a useful platform for expressing heterologous SSUs (Atkinson et al., 2017).

However, to enable co-engineering of SSUs and LSUs it would be desirable to develop a stable knock-out mutant for the major *rbcS* isoforms that contribute to Rubisco content in tobacco.

Unlike random mutagenesis techniques (e.g., T-DNA insertion and EMS) the targeted knockout of multiple genes (multiplex) can be achieved with relative ease using RGEN approaches. Several vector toolkits have been developed to clone multiple gRNAs into the same binary vector (Lowder et al., 2015; Ma et al., 2015; Xing et al., 2014). The majority of toolkits are compatible with popular modular cloning methods, including GoldenGate (Engler et al., 2014; Patron et al., 2015). Expressing multiple gRNAs with *Cas9* has produced homozygous sextuple mutants in Arabidopsis (Zhang et al., 2016) and octuple mutants in rice (Shen et al., 2017). An alternative strategy can be used for multiplex editing of homologous genes in polyploid genomes, or gene families with high sequence identity. In these cases, a "promiscuous" gRNA that targets a region of shared homology can be designed (Endo et al., 2015).

Multiplex targeting can also delete genomic regions between two gRNA targets located on the same chromosome (Ordon et al., 2017; Zhou et al., 2014). Dual gRNA approaches increase the likelihood of a frameshift mutation and have also been used to delete gene clusters (Zhou et al., 2014) and excise transgenes (Srivastava et al., 2017). Furthermore, creating a larger deletion can simplify screening for mutations in one or more genes. Indels are often detected by PCR and Sanger sequencing, restriction enzyme-based assays, and the T7E1 endonuclease (Xie and Yang, 2013) or SURVEYOR assays (Voytas, 2013). Large populations or multiple generations are needed to identify mutations in multiple genes owing to a high frequency of non-heritable mutations (Jansing et al., 2019) and screening can create a bottleneck for identifying multiplex mutant lines. In contrast, a deletion between two gRNA sites can allow for high-throughput screening as mutant and wild-type alleles can be distinguished directly by PCR (Gao et al., 2015; Ordon et al., 2017; Zhou et al., 2014). The ability to screen mutants by

PCR provides a more high-throughput method for identifying mutations and reduces the population of plants selected for sequencing of wild-type sized bands.

RGEN systems offer a strategy to edit multiple genes in complex polyploid species, including the *rbcS* family. This work aims to generate a stable tobacco line with reduced Rubisco content for use as a platform to express heterologous SSUs. Here, two major *rbcS* isoforms in tobacco, *rbcS-S1* and *rbcS-T1*, were knocked out using the RGEN *SpCas9*. Plants in the T_1 generation that harboured mutations in both isoforms and lacked the *SpCas9* transgene are described. Rubisco contents were reduced to *ca.* 5% of wild-type levels, which caused severe impairment of photosynthesis and growth. This work demonstrates the potential to manipulate *rbcS* isoforms using RGEN approaches in tobacco. Furthermore, the edited lines provide a tool for future engineering of the Rubisco enzyme *in planta*.

4.2 Results

4.2.1 Design of RNA-guided SpCas9 editing strategy

Two *rbcS* isoforms, *rbcS-T1* and *rbcS-S1,* were believed to account for the bulk of *rbcS* transcripts in tobacco leaves (Fig. 3–8). These isoforms share a high nucleotide sequence identity (*ca.* 95%), which allowed the design of promiscuous gRNA sequences that target both loci. Potential gRNA targets were selected that had a 20-nt region homologous to both isoforms and preceded by a PAM site (5'-NGG-3') for *SpCas9* recognition. Two suitable target sites in the first and fourth exons were identified that could result in 671-bp genomic deletion between the targets that would be distinguishable from wild-type sequences by PCR (Fig. 4–1A). The gRNA candidates were searched against the tobacco genome for potential off-target sites. It was not possible to design a gRNA without complementarity to any of the other seven *rbcS* isoforms due to high nucleotide sequence identity with *rbcS-T1* and *rbcS-S1* (*ca.* 88%). However, *SpCas9* is rarely active at target sites with mismatches in the gRNA sequence that occur in the seed region (11-12 bp adjacent to the PAM) of a target site (Peterson et al., 2016). Two gRNAs that had at least one mismatch in the seed region to potential off-target were chosen to target *rbcS-T1* and *rbcS-S1* (Fig. 4–1B).

Transient expression assays in tobacco were used to estimate the efficiency the gRNA pair, gRNA1-1 and gRNA1-2, at the two target sites, *rbcS-T1* and *rbcS-S1* (referred to as construct *SSU-T*) (Fig. 4–1A). The *spCas9* was constitutively expressed from the *Arabidopsis Ubiquitin-10* gene promoter (P_{UBQ10}) (Norris et al., 1993) and the gRNAs expressed from the *Arabidopsis U6* polymerase III promoter. P_{UBQ10} was selected to drive *SpCas9* expression due to high temporal stability in transient expression and increased frequencies of non-somatic mutations when *SpCas9* is stably expressed (Grefen et al., 2010; Khumsupan et al., 2019). Gene-specific primers (Table 2–1) spanning the target region were used to screen for mutations by PCR. If a deletion occurred between the two target sites a smaller (*ca*. 440 bp)

PCR amplicon would occur compared to wild-type (1.1 kb) (Fig. 4–1C). Alternatively, localised mutations induced at individual or both target sites could result in a small insertion or deletion (indel) that could be detected by sequencing the wild-type sized amplicon.

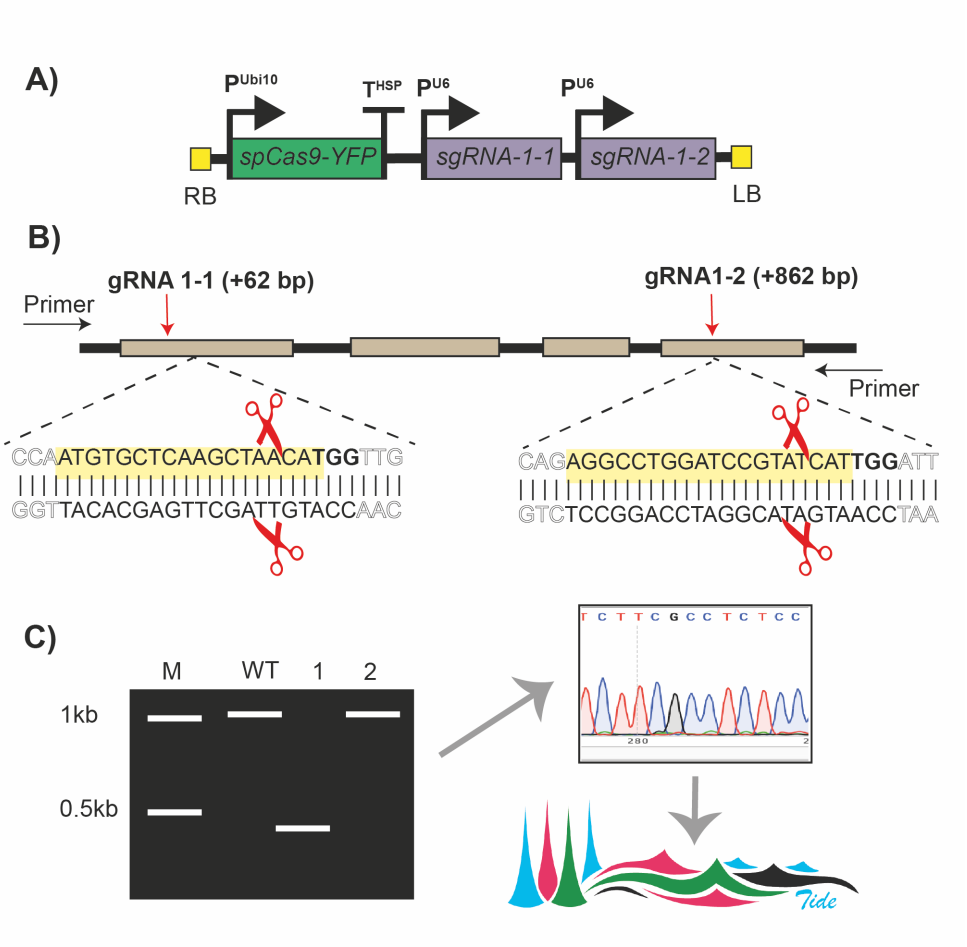


Figure 4-1. Genetic engineering strategy to generate knockout mutations in two major Rubisco small subunit isoforms in tobacco (rbcS-T1 and rbcS-S1). (A) Expression construct (*SSU-T*) used for transient expression assays. The *SpCas9* endonuclease with a nuclear-localisation signal (NLS) and yellow fluorescent protein (YFP) C-terminal tag expressed from a constitutive promoter P_{Ubi10}. Two guide RNAs (gRNAs) expressed from an RNA polymerase III dependent promoter (P_{U6}). **(B)** gRNA1-1 and gRNA1-2 target *spCas9* to exon 1 and exon 4 in *rbcS-T1* and *rbcS-S1*. A double-stranded break (DSB) is generated 3-4 bp upstream of the PAM motif (indicated in bold) at two independent sites **(C)** Expected deletion between the two target sites produces a 440 bp amplicon (lane 1) that is distinguishable from wild-type (WT) (1.1 kb) by PCR. Small insertions or deletions (indels) at one or both target sites produce a wild-type sized amplicon (lane 2). Indels are detected by Sanger sequencing and analysis using Tracking Indel DEcomposition (TIDE) software.

The *SSU-T* construct was delivered by agroinfiltration to the intact leaves of four-week-old tobacco plants. The expression of *SpCas9:YFP* was confirmed by imaging leaf discs 48 hours post-transfection (Fig. 4–2A). Genomic DNA (gDNA) was extracted from leaves transiently expressing *SpCas9:YFP* to identify mutation events. In non-transformed leaves, the wild-type sized amplicon was exclusively detected (Fig. 4–2B). A smaller amplicon for both target genes was predominantly amplified in samples from replicate infiltrated leaves. The wild-type sized amplicons were also detected for *rbcS-T1* and *rbcS-S1* in infiltrated leaves; however, at a lower intensity than in non-transformed leaves. Sequencing of the smaller fragments confirmed that a 671-bp deletion had occurred 3-4 bp upstream of the PAMs adjacent to the gRNA-1 and gRNA-2 sites (Fig. 4–2C). The mutation efficiency of a gRNA pair can be estimated by the intensity of the wild-type and mutated amplicon within a single PCR reaction (e.g., by densiometric analysis). However, the intensity of the wild-type band was not sufficient determine the mutation efficiency. It is possible that the wild-type allele was almost absent in infiltrated tissues (i.e. almost 100% efficiency) or that it was detected sufficiently because of PCR amplification bias of the more abundant smaller fragment (Kanagawa, 2003).

The *rbcS-T1* and *rbcS-S1* isoforms were previously determined to account for at least 60% of the total *rbcS* pool (Fig. 3–8). A major decrease in Rubisco (i.e., >40%) resulted in a reduction in chlorophyll content and a chlorotic leaf phenotype in Arabidopsis (Izumi et al., 2012). Thus, a frameshift mutation in both *rbcS-T1* and *rbcS-S1* could be expected to produce a chlorotic leaf phenotype; however, no chlorotic patches of cells were observed. A similar lack of phenotype resulting from a disruption of carotenoid biosynthesis was observed in transient assays to mutate the *NbPDS* gene in tobacco (Li et al., 2013) despite a photobleached phenotype in stable knockout lines. Because Rubisco is highly abundant and has a long half-life (Simpson et al., 1981), it is likely that the time of the assay was not sufficient to affect the accumulation of Rubisco. However, the transient assays demonstrate that the *SSU-T* construct can produce mutations in both isoforms with high efficiency.

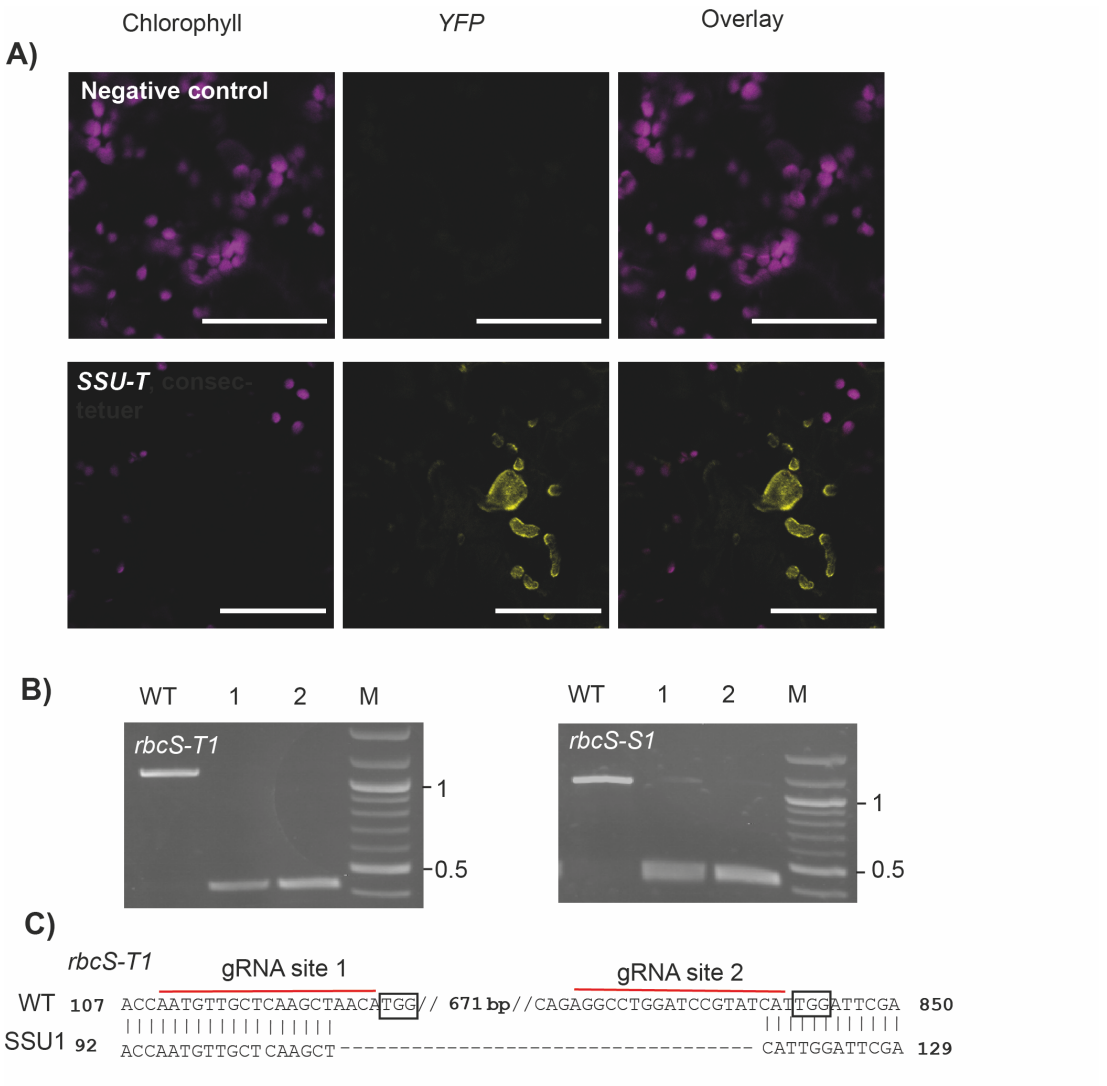


Figure 4-2. Transient expression of a Cas9-gRNA vector (*SSU-T*) targeting two *rbcS* isoforms in tobacco leaves. (A) Cas9:YFP expression is detected 48 h after agroinfiltration with *SSU-T* (bars = 50µm). **(B)** Two gRNAs targeted the two *rbcS* isoforms *rbcS-T1* and *rbcS-S1* and induced a 671 bp deletion in two biological replicates (lanes 1 and 2) that was distinguished from the wild-type amplicon (1.1 kb) by PCR using gene-specific primers. For *rbcS-S1,* the wild-type amplicon and the deletion band were detected in infiltrated leaves **(C)** Example of a pairwise sequence alignment between the wild-type amplicon of *rbcS-T1* (1.1 kb) and the suspected deletion band (lane 1) amplified from infiltrated leaves. Sanger sequencing confirmed a deletion event occurring between the gRNA 1-1 target site 4-bp upstream of the proximal adjacent motif (PAM) ('TGG') and 3-bp upstream of the gRNA 1-2 site.

4.2.2 Production of tobacco knockout line

Owing to little detection of the wild-type *rbcS-T1* and *rbcS-S1* amplicons in transient expression assays using *SSU-T*, a version of the construct without a YFP tag was used to generate stable SSU knock-out lines. It is essential to select mutated plants in the absence of the RGEN to distinguish between chimeric and bi-allelic mutations (Feng et al., 2014). Two different tissue culture approaches were explored to select transgene-free, edited plants. First, by the regeneration of tissue from plants transiently expressing a *Sp*Cas9 vector by agroinfiltration, and second, by stable integration of the transgene and segregation in subsequent generations (i.e. *Agrobacterium*-mediated transformation).

Agrobacterium-mediated transformation of leaf discs and subsequent regeneration of antibiotic and/or herbicide-resistant leaf disks provides a reliable and efficient method of gene transfer for many species (Horsch et al., 1989). Leaf discs or calli are typically co-cultivated with *Agrobacterium* before transfer to a selective media to regenerate shoots from cells with genomic integration events. Agroinfiltration of leaf discs (Jia et al., 2007) or whole leaves (Sparkes et al., 2006) is an alternative method of transformation. Although both approaches are sufficient to select for genomic integration events, agroinfiltration can increase the accessibility of plant cells to *Agrobacterium* and increase the potential transformation efficiency (Jia et al., 2007). *Nicotiana benthamiana* plants expressing *SpCas9* with mutations in the target locus were regenerated from agroinfiltrated leaves on a selective media (Nekrasov et al., 2013). The use of a selectable marker facilitates the selection of transformed plants by reducing the production of non-transformed shoots. However, in the case of gene editing, chromosomal integration of the RGEN transgene is not necessary to generate edited plants (Zhang et al., 2016). Furthermore, a key advantage of this approach includes the faster generation of lines with stable mutations as Cas9 does not need to be segregated out in subsequent generations.

Genomic integration of *SpCas9* and a transgene-free method were explored to produce plants with mutations in *rbcS-T1* and *rbcS-S1*. For the transgene-free method, tobacco leaves were agroinfiltrated with a marker-free version of *SSU-T* to examine if plants with mutations could be produced from cells transiently expressing unintegrated DNA. Leaves were surface sterilised five days post-infiltration as previously described (Sparkes et al., 2006) and 100 leaf discs were maintained on a non-selective shoot induction media. The regeneration efficiency was extremely low due to cell death arising from the sterilisation procedure. More than 90% of the leaf discs did not produce shoots, and the development of shoots from the remaining discs was severely impaired. Therefore, the transgene-free method was not pursued to produce transgene-free edited plants. For genomic integration of SpCas9 by *Agrobacterium*-mediated transformation, a version of *SSU-T* (*SSU1*) containing a kanamycin resistance marker (*nptII*) was transformed into leaf discs. Shoots that had the transgene integrated were selectively regenerated using kanamycin. A population of twenty regenerated plants was obtained to allow further analysis of mutations in the T_0 generation, and segregation of the transgene in the subsequent generation (T_1 generation).

4.2.3 Selection of knockout lines in the T_0 generation

Eight kanamycin resistant T_0 lines were selected and the integration of the *SpCas9* transgene was confirmed by PCR. Transgenic plants were then screened for mutation events in *rbcS-T1* and *rbcS-S1* by gene-specific PCR (Fig. 4–3B). A smaller amplicon was detected in *rbcS-T1* for a single plant line (*SSU-4*), indicating a deletion between the two gRNA sites. A wild-type sized fragment for *rbcS-T1* was observed for the seven other lines. All eight lines showed a wild-type sized fragment for *rbcS-S1*. All amplicons were sequenced to confirm the presence of mutations.

All eight plants expressing *SpCas9* contained two or more mutated alleles at one or both of the target sites for *rbcS-T1* (Table 4–1). Transgenic plants with two mutated alleles for a single

target site were putatively bi-allelic (*SSU-14*), while more than two alleles indicate somatic (chimeric) mutations (*SSU-1*, *SSU-2*, *SSU-3*, *SSU-9*, *SSU-12*, and *SSU-21*). Sequencing of the smaller amplicon for *rbcS-T1* in line *SSU-4* indicated a homozygous deletion of 671 bp, while line *SSU-14* showed putatively bi-allelic indel mutations at the second gRNA site for *rbcS-T1* (Fig. 4–3). The remaining six plants had chimeric mutations at the second gRNA site for *rbcS-T1*. The wild-type allele was only identified in chimeric plants and accounted for 15% of amplified sequences (Fig. 4–5). Two pairs of primers were used to amplify *rbcS-S1* from genomic DNA (Table 2-1). However, the control amplicon from wild-type leaves contained a mixture of *rbcS-S1*-like sequences that prevented analysis by pairwise sequence alignment and TIDE. Therefore, it was not possible to determine the mutation efficiency for *rbcS-S1* in the T_0 generation, possibly owing to amplification of both *rbcS-S1a* and *rbcS-S1b*. Mutations in *rbcS-S1* were suspected in five plants that had a range of chlorotic phenotypes (Fig. 4–4A) that were consistent with reduced Rubisco content (e.g. Arabidopsis *1a3b*).

Total soluble protein was extracted from five T_0 plants to estimate if a mutation in one or both isoforms decreased the amount of Rubisco (Fig. 4–4C). Immunoblotting of total soluble protein was performed to estimate the effect of the observed mutations in *rbcS-T1* on SSU content (Fig. 4–4C). Transgenic lines appeared to have lower amounts of Rubisco SSU in an equivalent leaf area compared to wild-type. Transgenic lines also had visible leaf phenotypes that were consistent with reduced Rubisco content (Fig. 4–4A). Plants with homozygous (*SSU-4*) or bi-allelic mutations (*SSU-14*) in *rbcS-T1* had pale and chlorotic leaves in comparison to wild-type. Chimeric plants had heterogeneous phenotypes of pale and wild-type cells, ranging from mostly pale (*SSU-9*) to mostly wild-type (*SSU-21*). The pale and wild-type phenotypes within the same leaf is in agreement with the presence of wild-type alleles detected in chimeric lines (Table 4–1). Two lines (*SSU-1* and *SSU-14*) did not establish on soil and are excluded from further analyses. PCR-based sequencing analysis of the T_0 generation indicates the

frequency of mutations but not the germ-line (heritable) mutation rate. Thus, the progeny of self-fertilised lines (T_1 generation) were screened to identify heritable mutations.

Table 4-1. Mutations and phenotypes of T_0 generation tobacco plants with mutations in *rbcS-T1*. Alleles were identified by Sanger sequencing of PCR products and analysis with Tracking Indels by DEcomposition (TIDE) software. Indels were detected exclusively at the second gRNA site in *rbcS-T1*. WT – wild-type, d – deletion, i – insertion followed by the number of base-pairs differing from the WT allele.

Line	Phenotype	Zygosity	Alleles
SSU-1	Pale	Chimeric	d6, d3, d1, i1
SSU-2	Chimeric	Chimeric	d9, d6, d5, d1
SSU-3	Chimeric	Chimeric	WT, d6, d4, d3, d2, d1, i1
SSU-4	Pale	Homozygous	d670
SSU-9	Pale	Chimeric	d17, d3, d1
SSU-12	Chimeric	Chimeric	WT, d15, d13, d6, d1
SSU-14	Pale	Bi-allelic	d1, i1
SSU-21	Chimeric	Chimeric	WT, d5, d3, d2, i1

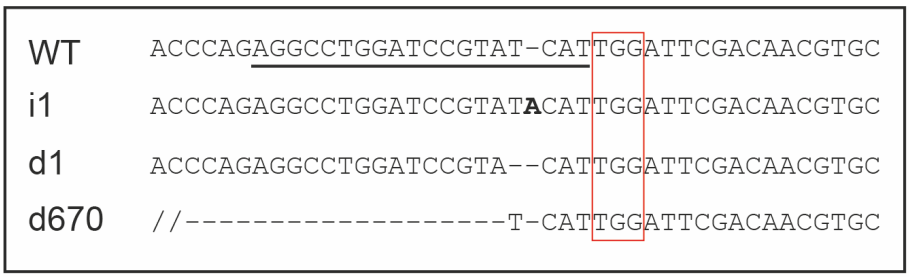


***Figure 4-3*. rbcS-T1 *alleles* identified in two suspected non-chimeric *SSU* T0 plants.** Alignments show a 1-bp insertion (i1) and 1-bp deletion (d1) at the second gRNA site (underlined) identified in a biallelic plant (SSU-14) and a 670-bp homozygous deletion (*SSU-4*) (Appendix B Fig. 3 shows the full alignment for *SSU-4*) detected as described in Table 4-1. The red box shows the PAM motif (5'-TGG-3').

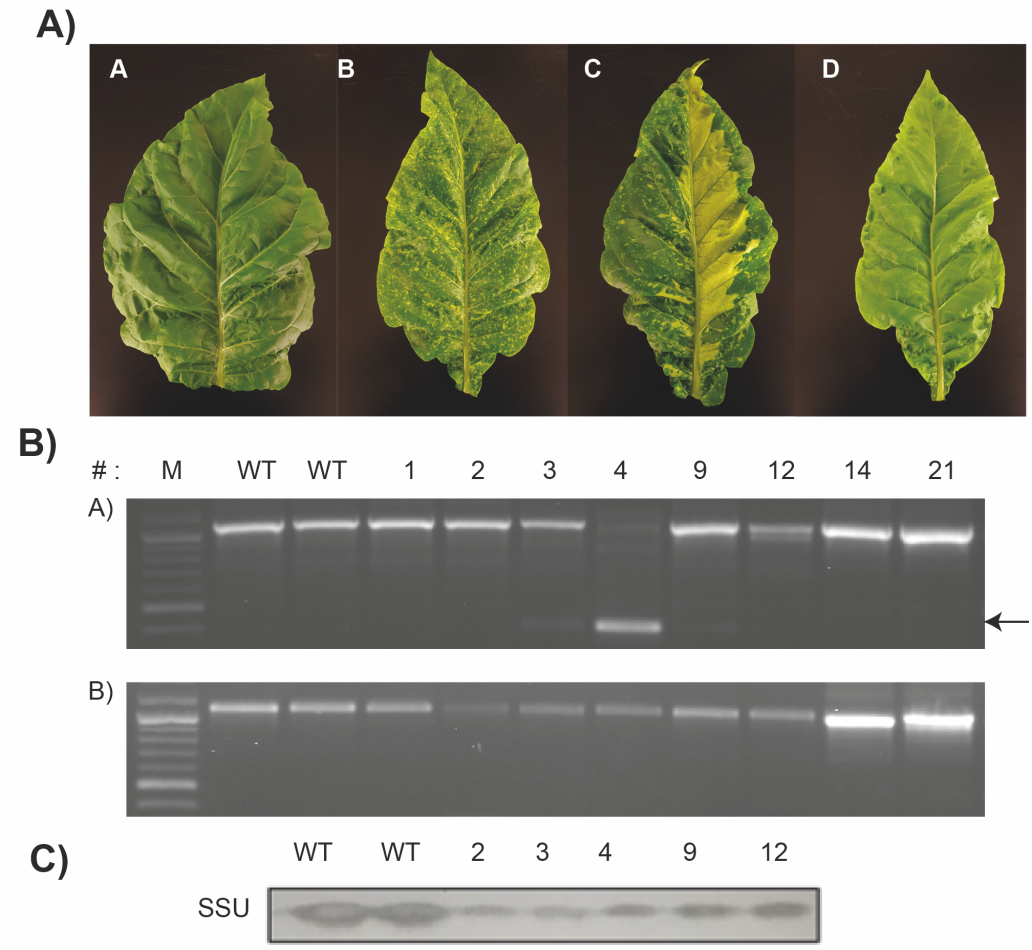


Figure 4-4. Pale leaf phenotype of transgenic *SSU* plants in the T_0 generation and mutation screening by PCR. (A) Examples of the range of phenotypes produced from *spCas9* mediated targeting of two *rbcS* isoforms (A) wild-type (WT) tissue culture control. (B, C) chimeric mutants with mixed pale and wild-type phenotypes. (D) homozygous and/or bi-allelic mutant. **(B)** Gene-specific primers amplify a 1.1 kb fragment in wild-type plant and mutants with chimeric, heterozygous, bi-allelic or homozygous indels at one or both target sites. **(B)** A deletion between the two gRNA sites in *rbcS-T1* produces a 460 bp amplicon (indicated by an arrow). B) Primers for *rbcS-S1* also produced a 1.1 kb amplicon for all lines. M = DNA marker. **(C)** Immunoblotting of total soluble protein extracts to detect the 14.5 kDa Rubisco small subunit (SSU)

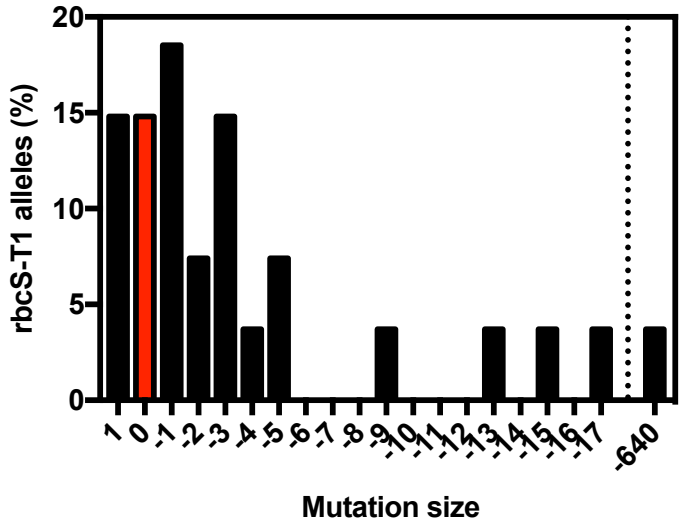


Figure 4-5. Frequency of wild-type (0) and mutated alleles in *rbcS-T1* detected in eight plants expressing Cas9-gRNA. Different alleles were detected by analysing sequencing chromatograms using TIDE software. Analysis of each chromatogram identified multiple alleles (including the wild-type allele) in chimeric plants and one or two alleles for homozygous and bi-allelic plants respectively. Data shows all of the alleles identified as a percentage of total alleles (Table 4-1). Mutation size describes the number of base pairs inserted or deleted compared to the wild-type allele.

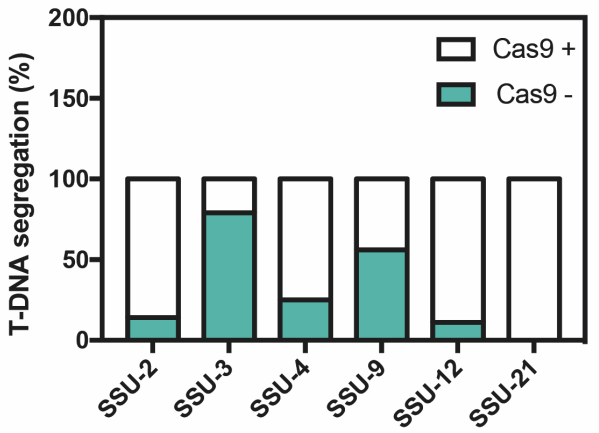


Figure 4-6. Segregation of *SpCas9* T-DNA in the T_1 progeny of six tobacco lines. Transgenic (Cas9 +) and transgene-free (Cas9 -) plants were identified by gene-specific PCR (n=20 plants per line).

4.2.4 Heritability of mutations in the T_1 generation

Six T_1 lines (*SSU-2, SSU-3, SSU-4, SSU-9, SSU-12,* and *SSU-21*) were germinated on soil and screened for the Cas9 transgene by PCR. Plants were not germinated on a selective media to identify heritable mutations in transgene-free lines. Transgene-free plants were identified in all lines except *SSU-21* (Fig. 4–6). One line (*SSU-4)* had a 3:1 segregation ratio for a single-copy insertion. Transgenic and transgene-free plants were subsequently screened by gene-specific PCR to determine the zygosity of mutations at the target sites.

Five of the six lines studied had putative chimeric mutations in the T_0 generation (Table 4-1). Although chimeric plants can produce progeny with stable mutations, the transmission of alleles is unpredictable and new mutations are likely to arise (Feng et al., 2014; Pan et al., 2016). Similarly, bi-allelic plants can have unexpected patterns of inheritance, with the unequal inheritance of alleles occurring (Pan et al., 2016; Xu et al., 2015). Plants with mutations in *rbcS-T1* were identified in five out of six lines (*SSU-2, SSU-3, SSU-4, SSU-9*, and *SSU-12*) (Table 4-2). The progeny of *SSU-4* were homozygous for a 640 bp deletion between the two gRNA sites. The five chimeric lines *SSU-2*, *SSU-3*, *SSU-9, SSU-12,* and *SSU-21*, had variable frequencies of mutations in the T_1 generation. No wild-type alleles were detected in the progeny of *SSU-2, SSU*-9, and *SSU-12*. Three transgene-free *SSU-2* plants had bi-allelic mutations, while the remaining five transgenic plants were chimeric. Two different bi-allelic mutations were identified in line *SSU-2*: two plants had the genotype -5/-1 and the third plant was -9/-6. Four *SSU-9* plants had a homozygous deletion (1-bp) and two had different bi-allelic mutations. One *SSU-12* plant was homozygous (1-bp deletion) and four were putatively bi-allelic due to the presence of the transgene. Heterozygous mutations with the wild-type allele were only detected in line *SSU-3* and accounted for the majority of transgene-free genotypes. The progeny of one line (*SSU-21*) had a visible chimeric phenotype and was not screened for mutations.

Previous attempts to screen for mutations in *rbcS-S1* were not successful. However, it was possible to specifically amplify *rbcS-S1* using a cDNA template. Three lines with homozygous or bi-allelic mutations in *rbcS-T1* (*SSU-2, SSU-4, SSU-9, SSU-12*) were screened for mutations in *rbcS-S1* and the sequencing chromatograms were analysed by TIDE. Two types of homozygous mutations were identified in three *SSU4* plants and the remaining progeny (7/10 plants) were putatively bi-allelic. Bi-allelic mutations were also identified in half of the progeny of *SSU-9* (4/9 plants). The remaining *SSU-9* progeny were suspected chimeras due to the presence of more than two alleles at the second gRNA site. All of the progeny of *SSU-12* (10 plants) had more than two alleles for *rbcS-S1*; however, all of the detected alleles were mutated and the wild-type allele was not identified.

Three distinct phenotypes were observed in transgenic and non-transgenic T_1 plants. Pale leaves and stunted growth were observed in non-transgenic homozygous and bi-allelic plants for lines *SSU-2, SSU-4, SSU*-9, and *SSU-12*. Transgenic plants with chimeric mutations and lacking the wild-type allele were observed for lines *SSU-2, SSU-3, SSU-9,* and *SSU-12*, which had the same phenotype as plants with heritable mutations. Chimeric plants harbouring the wild-type allele (*SSU-21*) had variegated leaves, while non-transgenic heterozygous plants (*SSU-3*) had a wild-type-like phenotype.

Heritable mutations in *rbcS-T1* and *rbcS-S1* were identified in the T_1 generation of five out of six lines. Although segregation of the transgene was lower than expected, transgenic plants lacking the wild-type allele had phenotypes consistent with homozygous or bi-allelic mutations. A mixture of transgenic and non-transgenic plants from three lines (SSU-4, SSU-9 and SSU-12) that lacked the wild-type alleles were selected to further examine the effect of the mutated *rbcS-T1* and *rbcS-S1* on photosynthesis and growth.

Table 4-2. Inheritance of *SpCas9* mutations in two Rubisco small subunit genes (*rbcS-S1* and *rbcS-T1*) in the T_1 generation. The number of T_1 progeny with homozygous (hom), heterozygous (het), bi-allelic (bi) or chimeric (chi) mutations is shown. Non-chimeric alleles in the T_1 generation are described as deletions (d) or insertions (i) followed by the number of base-pairs compared to wild-type. Alleles that were not identified in the T_0 progenitor are shown in bold lettering. The number of progeny with a single genotype is subsequently shown in brackets.

Line	Isoform	Zygosity (T_0)	Zygosity (T_1)	Alleles (T_1)
SSU-2				
	rbcS-T1	Chimeric	3bi;5chi	d1/d5 (2)
				d9/d6 (1)
	rbcS-S1	n/a	n/a	
SSU-3				
	rbcS-T1	Chimeric	1bi;4het;10chi	d2/i1 (1)
				d2/WT (3)
				i1/WT (1)
	rbcS-S1	n/a	n/a	
SSU-4				
	rbcS-T1	Homozygous	16hom	d670 (16)
	rbcS-S1	n/a	3 hom;7bi	d1(1)
				d2 (2)
				d1/d2 (7)
SSU-9				
	rbcS-T1	Chimeric	2hom;2bi;4chi	d1 (2)
				d17/d3 (1)
				d1/**d6** (1)
	rbcS-S1	n/a	4bi;5chi	d5/d3 (4)
SSU-12				
	rbcS-T1	Chimeric	6bi; 6chi	**d2**/**d7** (2)
				d1/d6 (1)
				d9/**d4** (3)
	rbcS-S1	n/a	10 chi	
SSU-21				
	rbcS-T1	Chimeric	n/a	n/a
	rbcS-S1	n/a	n/a	n/a

4.2.5 Molecular and biochemical analysis of SSU lines

Twelve T_1 plants from each independent line (*SSU-4*, *SSU-9* and *SSU-12*) were grown in a greenhouse for an additional two weeks alongside a non-transformed wild-type control (WT). A range of mutation types was identified within each line; however, the selected lines were considered a unified population as the wild-type allele was not detected. Subsequently, the amount of *rbcS* and *rbcL* transcripts and Rubisco content of these three lines were investigated using qPCR and ^{14}C-CABP binding respectively.

The abundance of *rbcS-S1* and *rbcS-T1* transcripts was significantly lower in *SSU-4,* while lines *SSU-9* and *SSU-12* maintained wild-type levels (Fig. 4–7). The absence of detectable *rbcS-T1* transcripts in *SSU-4* plants was likely because of the 670-bp deletion event (Table 4-2), which removed the forward primer binding site for the qPCR assay. Expression of the third most abundant isoform (*rbcS-S5*) and two minor isoforms (*rbcS-T2* and *rbcS*-T4) did not differ from wild-type. Three of the minor isoforms (*rbcS-S2*, *rbcS-S4*, and *rbcS-T5*) were upregulated in the *SSU* knockout lines and *rbcS-S3* expression was increased in *SSU-12*. As the three upregulated isoforms account for *ca.* 12% of *rbcS* transcripts, the increase in expression did not significantly affect the total transcript level (Fig. 4–8). In contrast, the LSU-encoding *rbcL* transcripts decreased to *ca.* 80% of wild-type levels in all *SSU* knockout lines (Fig. 4–8).

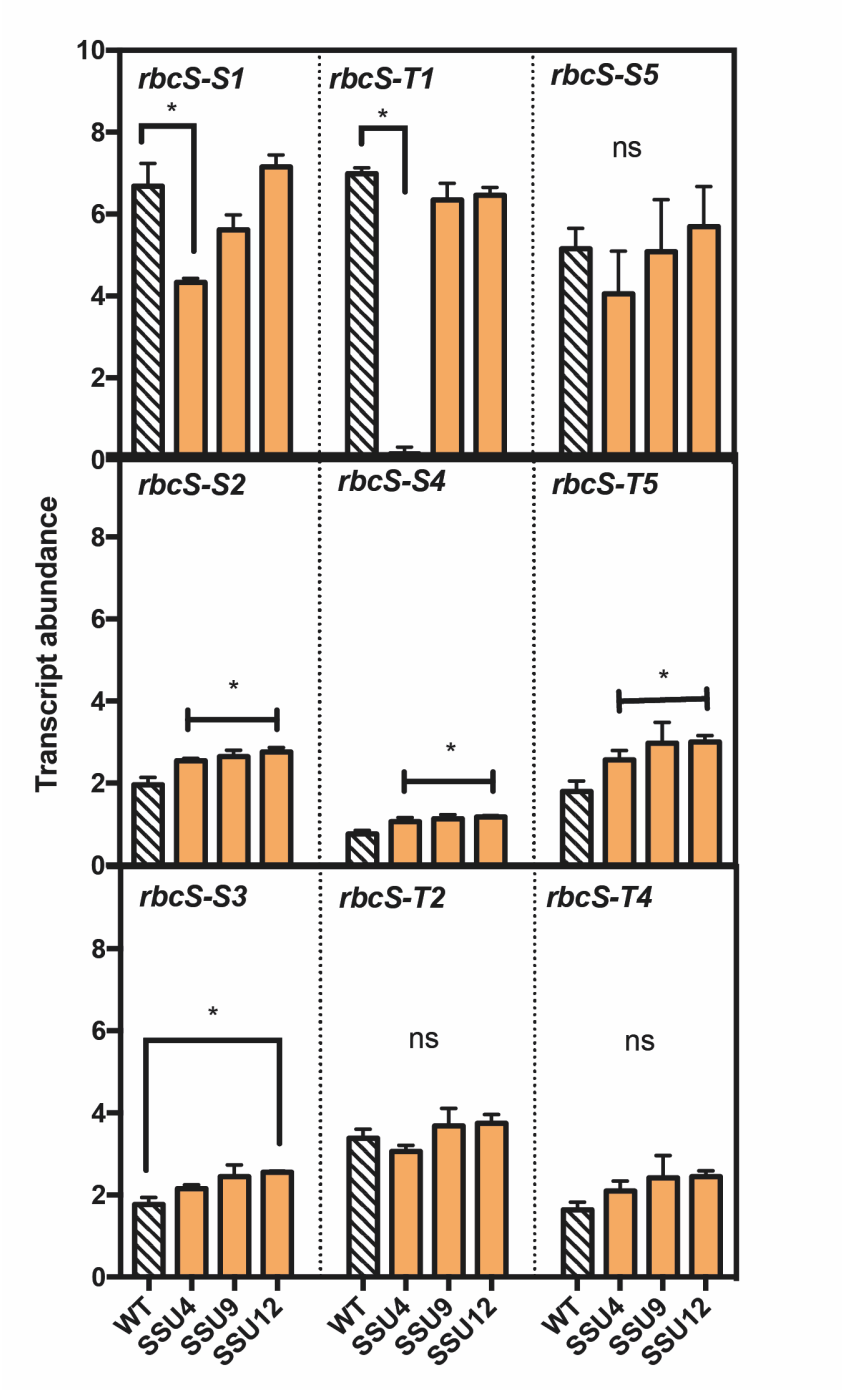


Figure 4-7. Expression of individual rbcS isoform transcripts in lines with spCas9 induced mutations in rbcS-T1 and rbcS-S1. RNA was extracted from wild-type and plants in the T_1 generation from three independent lines (SSU-4, SSU-9 and SSU-12). Quantitative PCR (qPCR) was used to determine the C_T value of rbcS transcripts, along with a constitutively expressed housekeeping gene (L25). Transcript abundance was determined by normalising the target C_T value to the housekeeping C_T value using the ΔCT method. (n = 4). Significant differences ($p<0.001$) to wild-type are shown (*), ns = no significance.

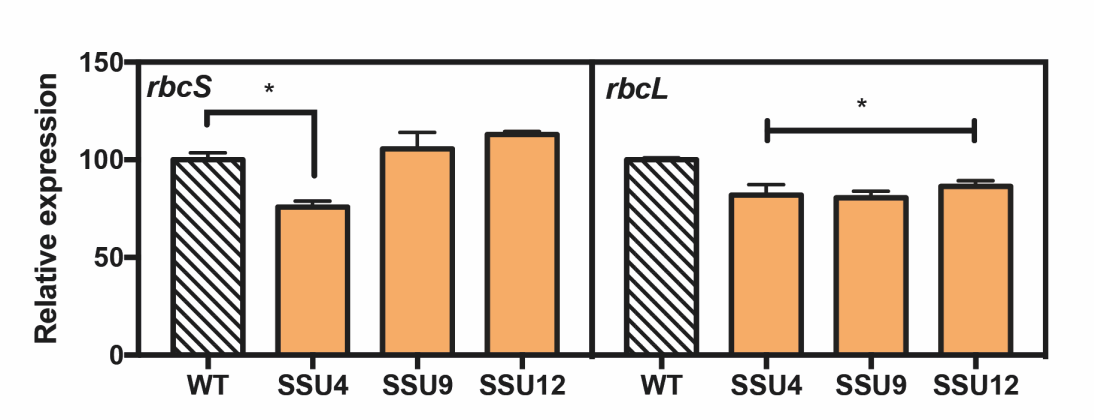


Figure 4-8. The abundance of Rubisco small subunit (rbcS) and large subunit (rbcL) transcripts in lines with spCas9 induced mutations in two rbcS isoforms. Normalised C_T values for total rbcS transcripts (nine isoforms) and rbcL transcripts are shown for three independent lines (SSU-4, SSU-9 and SSU-12) are shown relative to wild-type (WT).

Although the mutations in *rbcS-1* and *rbcS-T1* did not severely affect the amount of *rbcS* and *rbcL* transcripts, the Rubisco content was reduced to *ca.* 7% of wild-type levels (Fig. 4–9A). Decreased Rubisco was associated with *ca.* 85% less SSU and 60% and LSU as estimated by immunoblotting (Fig. 4–9C). The reduced accumulation of LSU peptides suggests that the mutations in *rbcS-1* and *rbcS-T1* affected the production or stability of the LSU. All three mutant lines also had 70-80% less total soluble protein that was not accounted for solely by decreased Rubisco (Fig. 4–9B). Wild-type had a 3 g m^{-2} investment in Rubisco, accounting for 30-40% of TSP while TSP reductions in the mutant were greater indicating reduced the synthesis of proteins other than Rubisco. *SSU* lines also had significantly less chlorophyll per leaf area compared to wild-type (Fig. 4–10). Total chlorophyll was reduced to 30-40 % of wild-type levels, while no changes were observed in the ratio of chlorophyll a/b.

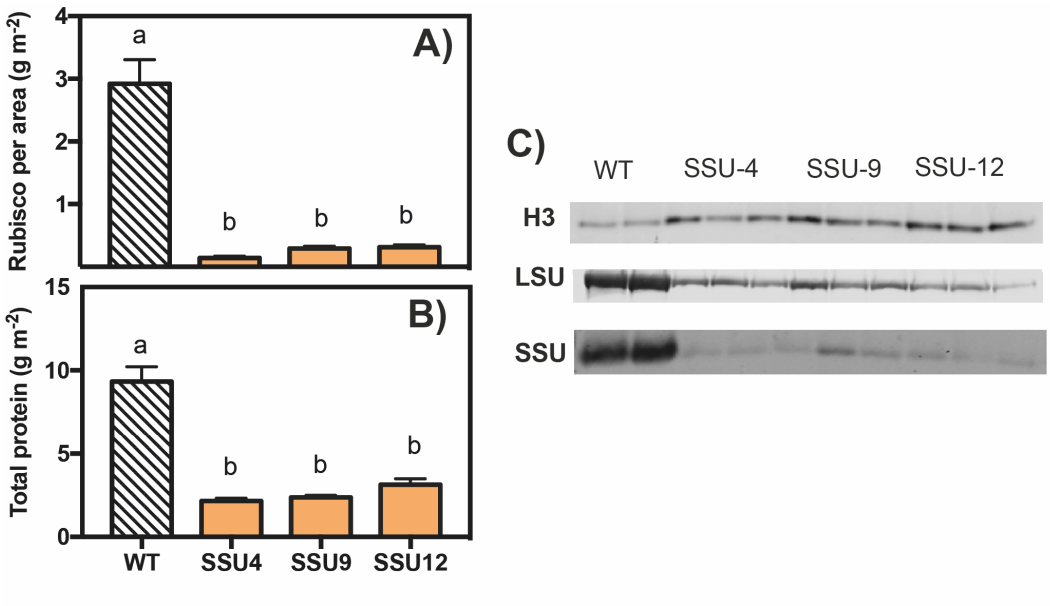


Figure 4-9. Molecular and biochemical analysis of plants with mutations in two Rubisco small subunit (SSU) isoforms (*rbcS-T1* and *rbcS-S1*). (A) Rubisco content per leaf area determined for wild-type (WT) and three independent lines (*SSU-4*, *SSU-9* and *SSU-12*). **(B)** Total protein per leaf area determined by Bradford assay. **(C)** Immunoblot of total protein to detect the LSU, SSU, and histone 3 (H3) loading control. Protein extracted from an equivalent amounts of leaf area was loaded. Each lane represents a biological replicate (2 for WT and 3 for each SSU line).

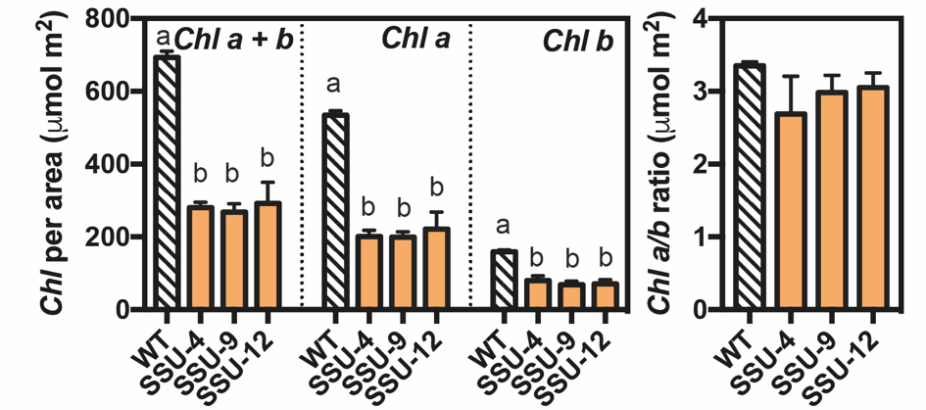


Figure 4-10. Chlorophyll content in plants with reduced Rubisco content. The amount of chlorophyll a and chlorophyll b was determined in equivalent leaf area in wild-type and three SSU mutant lines. Values represent the mean ±SEM (n=4) followed by different letters indicating significant differences determined by ANOVA followed by Tukey's HSD tests ($P<0.05$)

5.2.6 Photosynthesis of SSU lines

The response of CO_2 assimilation (A) to photosynthetically active radiation (PAR) was measured under ambient CO_2 concentrations (Fig. 4–11A). At low light (≤ 100 μmol photons m^{-2} s^{-1}) the initial slope of the photosynthetic response was equivalent between wild-type and SSU mutants. However, the SSU plants had a significantly lower light-saturated rate of CO_2 assimilation (A_{max}) (*ca.* 42% of wild-type) that occurred at a moderate light intensity (400 μmol photons m^{-2} s^{-1} compared to 1000 μmol photons m^{-2} s^{-1} for wild-type plants) (Table 4–3). Under saturating light (1500 μmol photons m^{-2} s^{-1}), the response of photosynthesis *(A*) to changes in the internal CO_2 concentration (C_i) was also affected in SSU mutants (Fig. 4–11B). The linear part of the A/C_i response curve is limited by the carboxylation efficiency of Rubisco and was significantly lower in *SSU* plants than wild-type plants (Table 4–3). Therefore, A was decreased at lower CO_2 concentrations in the SSU lines than in wild-type plants. Furthermore, RuBP regeneration limits the rate at which A is saturated (Farquhar et al., 1980; Long and Bernacchi, 2003; Sharkey, 1985). The SSU lines did not reach the limit of RuBP regeneration capacity, even at high C_i levels; therefore, these lines had a severe impairment of photosynthesis.

The A/C_i curves were used to derive additional photosynthetic parameters to provide an understanding of the mechanisms underlying the reduction in A for the three mutant lines (Table 4–3). The CO_2 compensation point (Γ) was significantly higher and the maximum rate of carboxylation by Rubisco ($V_{c,max}$) was decreased by more than 50% compared to wild-type, indicating that A was Rubisco-limited in SSU mutants. Similarly, the maximum rate of electron transport (J_{max}) decreased by more than half compared to wild-type. The maximum quantum efficiency of PSII (F_v/F_m) was also significantly lower in SSU mutants. The ratio of F_v/F_m is consistently *ca.* 0.83 in non-stressed leaves (Baker, 2008). Therefore, a decrease in F_v/F_m to <0.8 in SSU mutants but not wild-type plants indicates stress in the *SSU* lines.

Table 4-3. Photosynthetic parameters of tobacco plants with reduced Rubisco contents derived from gas exchange analysis. The *A/Q* response curves were used to derive the maximum rate of photosynthesis (A_{max}) (Fig. 4–11A). The maximum rate of carboxylation by Rubisco ($V_{c,max}$), maximum electron transport rate (J_{max}) and CO_2 compensation point (Γ) were estimated by fitting the A/Ci response curves (Fig. 4–11B). Values represent the mean ±SEM (n=4). Dark-adapted leaves were used for F_v/F_m measurements (n=10). Different letters indicate significant differences determined by ANOVA followed by Tukey's HSD tests ($P<0.05$). Abbreviations: WT, wild-type.

	WT	***SSU-4***	***SSU-9***	***SSU-12***
A_{max} (μmol CO_2 m^{-2} s^{-1})	21.6 ± 0.8[a]	9.3 ± 0.3[b]	9.4 ± 0.3[b]	9.8 ± 0.4[b]
$V_{c,max}$ (μmol CO_2 m^{-2} s^{-1})	98.9 ± 1.2[a]	41.2 ± 5.4[b]	48.7 ± 2.2[b]	43.8 ± 7.5[b]
J_{max} (mmol e^- m^{-2} s^{-1})	187 ± 1.4[a]	73.0 ± 6.7[b]	67.4 ± 0.7[b]	71.0 ± 1.4[b]
Γ (μmol CO_2 mol^{-1})	53.0 ± 0.5[a]	96.7 ± 1.8[b]	92.1 ± 1.8[b]	95.7 ± 2.9[b]
Initial slope (A/C_i)	0.137 ±0.004[a]	0.027±0.005[b]	0.026±0.004[b]	0.030±0.001[b]
F_v/F_m	0.85 ± 0.03[a]	0.76 ± 0.02[b]	0.67 ± 0.02[c]	0.77 ± 0.02[b]

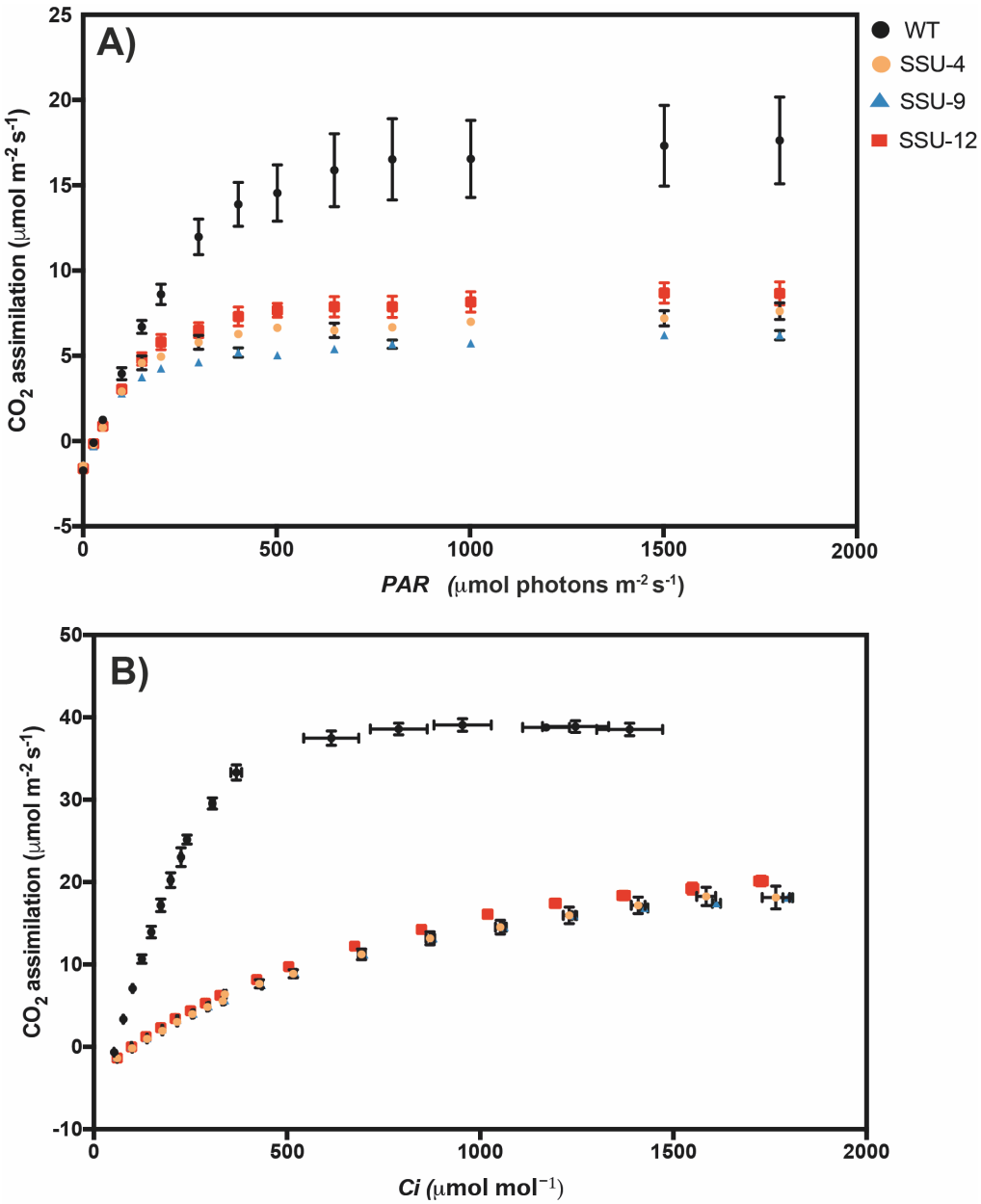


Figure 4-11. Photosynthetic response curves of tobacco Rubisco SSU mutants. (A) The response of *A* to changing irradiance (*Q*) measured at 25°C under atmospheric CO_2 concentrations (400 µmol mol^{-1}). Each data point represents the mean ± SEM of four leaves each from a separate plant (n = 4). **(B)** The response of *A* to intracellular CO_2 concentration (C_i) measured at 25°C under saturating irradiance (1500 µmol photons m^{-2} s^{-1}).

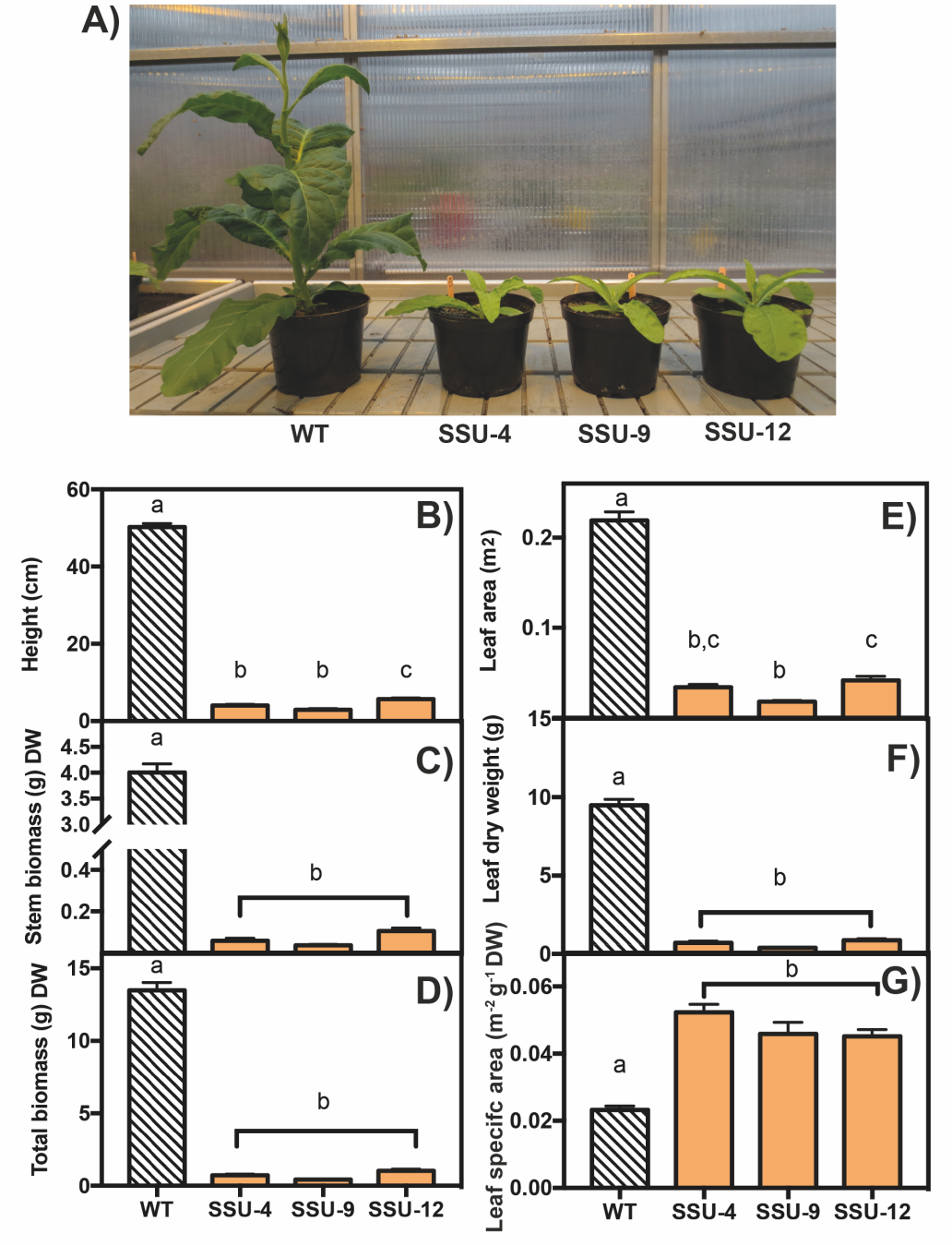


Figure 4-12. **Growth of tobacco Rubisco SSU mutants in the T_1 generation. (A)** Phenotypes of wild-type (WT) and three independent lines. **(B)** Stem height. **(C)** Stem dry weight. **(D)** Total dry weight. **(E)** Total leaf area. **(F)** Leaf dry weight. **(G)** Leaf area per plant dry weight. Bars represent the mean ±SEM (n =6-8) with significant differences ($p<0.05$) indicated by different letters determined by ANOVA and Tukey's HSD tests. Plants were grown in a greenhouse with a 14 h photoperiod under natural light data shows measurements for 45-day old plants.

4.2.7 Growth phenotypes of Rubisco SSU mutant lines

The three Rubisco SSU mutant lines had significantly reduced growth compared to wild-type plants after 45 days of growth (Fig 4–12A), and accumulated *ca.* 10% of wild-type total biomass (Fig. 4–12D). Mutant lines were shorter than wild-type (Fig. 4–12B) and stem biomass was reduced by *ca.* 99% (Fig. 4–12C). Total leaf area and leaf biomass were also reduced (Fig. 4–12E, 12F), but the number of leaves was not significantly different. The ratio of leaf fresh weight to dry weight (FW/DW) was at least twice that of wild-type leaves. Similarly, the leaf specific area (leaf area per dry weight) of the SSU mutants was increased by 50% compared to wild-type leaves (Fig. 4–12G).

4.2.8 Effect of *SpCas9* transgene on growth and photosynthesis

A mixed population of SSU mutants with or without the *SpCas9* transgene was used for the growth analysis. (Figure 4–13A). In summary, 50% of mutants from lines SSU-4 and SSU-9 were transgene-free, while one plant from line SSU-12 lacked the transgene. An additional analysis was performed on the results of the growth and photosynthesis experiments to determine if the presence of the *SpCas9* transgene affected the growth of SSU mutants (Fig 4–13B-13G). Transgenic plants had no significant differences in stem growth (Fig. 4–13B, 13C) or leaf biomass and area (Fig. 4–13E, 13F) compared to transgene-free plants. Similarly, the response of light-saturated photosynthesis to intracellular CO_2 concentrations was equivalent between the two populations of plants (Fig. 4–14).

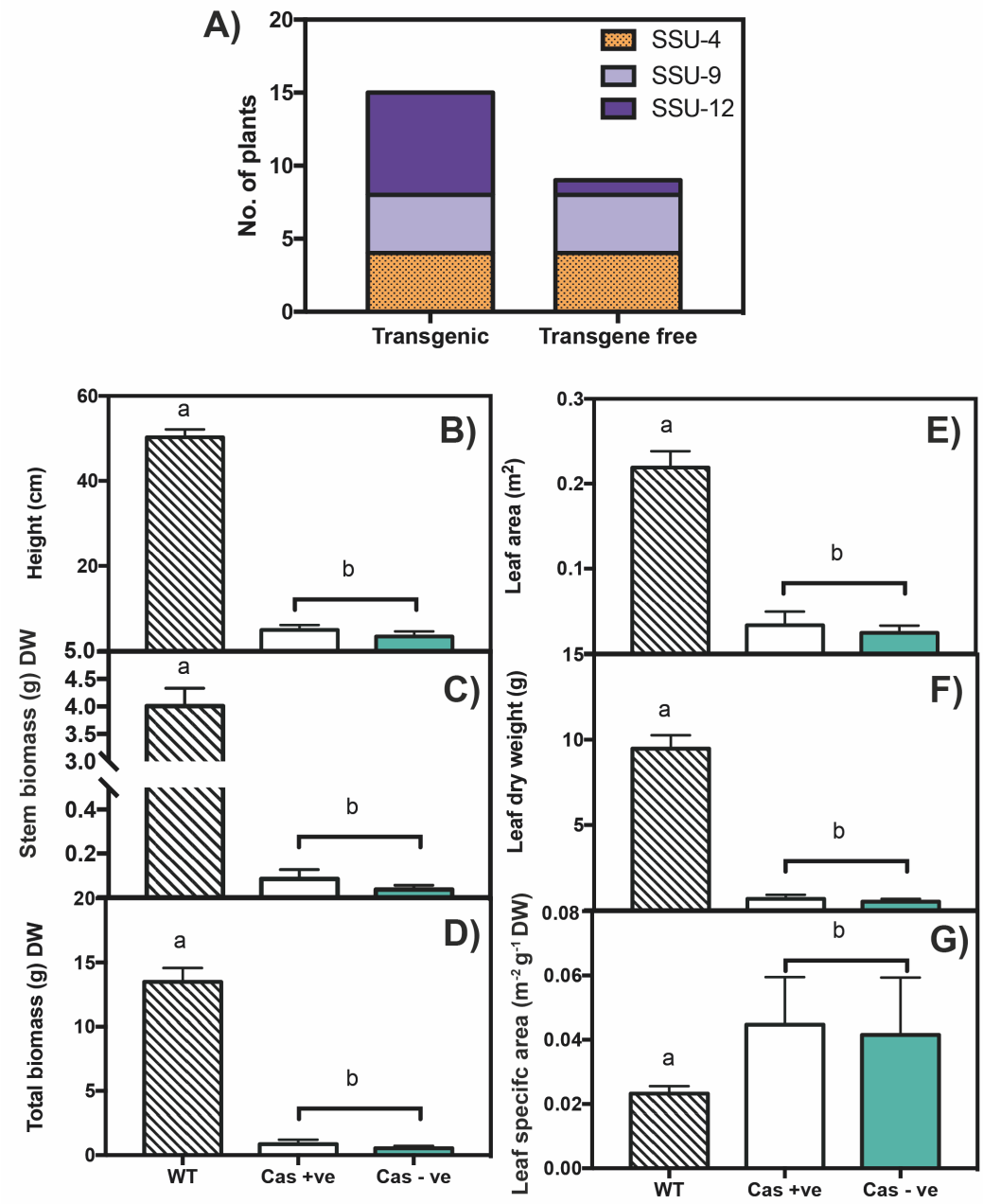


Figure 4-13. Analysis of growth parameters for transgenic and transgene-free plants with mutations in two *rbcS* isoforms. (A) Plants in the T_1 generation from three lines were mixed populations that had segregated (transgene free) or inherited the *spCas9* T-DNA. **(B)** Stem height. **(C)** Stem dry weight. **(D)** Total dry weight. **(E)** Total leaf area. **(F)** Leaf dry weight. **(G)** Total leaf area per unit of leaf dry weight. Bars represent the mean ±SEM (n = 10-15 with significant differences ($p<0.05$) indicated by different letters determined by ANOVA and Tukey's HSD tests.

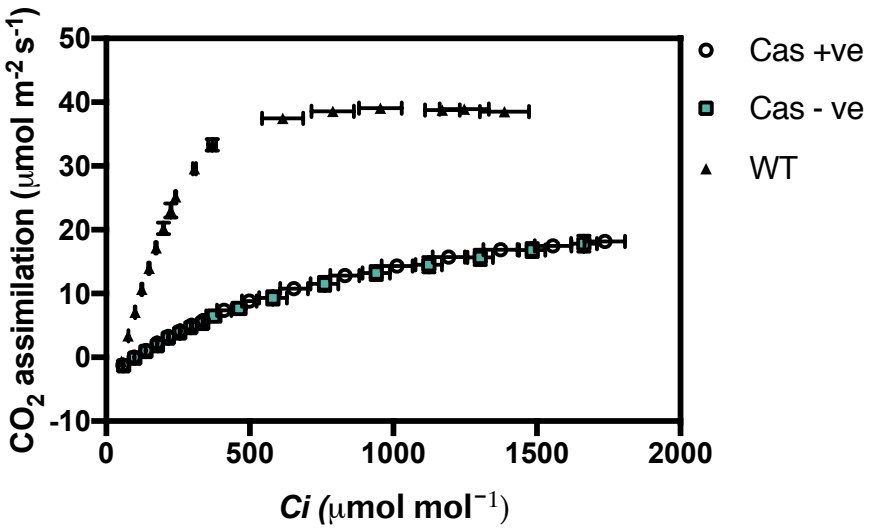


Figure 4-14. Analysis of photosynthetic response in transgenic (Cas +ve) and transgene-free (Cas -ve) plants with reduced Rubisco content. Measurements were as described for Fig. 4–12B

4.3 Discussion

Tobacco mutants with decreased amounts of Rubisco were generated by knocking out two *rbcS* isoforms (*rbcS-T1* and *rbcS-S1*). Gene knockout using the RGEN *SpCas9* generated loss-of-function mutations for both *rbcS* isoforms in T_1 generation plants that lacked the *SpCas9* transgene. This work demonstrates the use of RGENs for manipulating *rbcS* isoforms *in planta*. Furthermore, the knockdown *SSU* lines have a similar reduction in Rubisco content, as antisense tobacco lines with reduced amounts of Rubisco (Quick et al., 1991c).

Six transgenic lines were obtained with mutations in *rbcS-T1* and *rbcS-S1*. The mutation efficiency for *rbcS-T1* (85%) in the T_1 generation was similar to the reported rate of gene editing in tobacco using *SpCas9* (*ca.* 80%) (Gao et al., 2015). Mutation efficiencies can vary depending on the gRNA sequence and the target loci (Doll et al., 2019; Zhou et al., 2014). Due to high variability between targets, it is useful to test different gRNA candidates by transient

expression of RGEN vectors. Transient expression assays in protoplasts and leaves suggested that a large deletion between the target sites occurred with high efficiency. Deletion of the region between the two gRNA sites in T_0 plants occurred at a frequency of 12.5% (1/8 plants) in *rbcS-T1* but was not detected in *rbcS-S1*. Dual gRNAs have been used to delete large chromosomal regions (170 – 245 kb) in rice protoplasts (Zhou et al., 2014), but chromosomal deletions were detected at lower frequencies in transformed calli (6.25% and 16% respectively). Although deletions between two DSB sites with dual gRNAs have been reported in Arabidopsis, *Nicotiana benthamiana,* and *Zea maize*, indels at one or both target sites are more commonly detected (Doll et al., 2019; Durr et al., 2018; Ordon et al., 2017). Nevertheless, a high frequency of genomic deletions using dual gRNAs has been reported in Arabidopsis (Zhang et al., 2017; Zhao et al., 2016). Four independent loci were targeted in separate transformation events and deletions ranging from 255-934 bp were detected at a frequency of 20%-24%. Differences in the mutation efficiency of each gRNA can reduce the frequency of deletion between the target sites. A comprehensive analysis of the outcomes of dual gRNA strategies in *Zea mays* using two target sites located 100 bp apart reported that 78% of mutations did not involve deletions between the two targets. Instead, a clear preference for indels at one of the target sites (74%) was reported. Similarly, higher editing efficiency was found at the second gRNA site in *rbcS-T1*. A large difference in the editing efficiency between the two gRNAs could account for the low frequency of deletions between the target sites. Therefore, large deletions are feasible in *planta,* but the higher abundance of indels limits the use of the paired gRNA approach to facilitate screening. Generating smaller (50-100 bp) deletions is likely a more suitable strategy to create mutations that can be detected by PCR (Ordon et al., 2017).

Few studies have reported RGEN-mediated editing in tobacco (Gao et al., 2015; Xie et al., 2017) and the germline transmission rate of mutations has not yet been described. The generation of homozygous, heterozygous, and bi-allelic mutations in the first generation (i.e.,

T_0) has been reported in several species, including tomato (*Solanum lycopersicum*) and rice (*Oryza sativa*) (Brooks et al., 2014; Zhang et al., 2014); however, somatic mutations are more frequently detected. The frequency of non-somatic mutations varies depending on the target site and between transgenic events (Zhang et al., 2014). Somatic mutations occur after the single-cell stage of embryogenesis. However, it is possible for chimeric plants to transmit homozygous, heterozygous, or bi-allelic mutations in germline cells to the next generation (Feng et al., 2014; Zhang et al., 2014). Here, the segregation patterns of mutations in the progeny of chimeric *SSU* T_0 plants were complex and varied between independent lines. For example, the progeny of *SSU-21* had wild-type or chimeric genotypes, suggesting that *SpCas9* was not active in the germ-line cells of the T_0 progenitor. Bi-allelic mutations were identified in the progeny of two chimeric T_0 lines (*SSU-9* and *SSU-12*) at different frequencies. Half of the T_1 progeny of *SSU-9* were bi-allelic compared to *ca.* 25% of the progeny of *SSU-12*. The remaining progeny of *SSU-9* and *SSU*-12 lacked the wild-type allele but because more than two mutated alleles were detected they were putatively chimeric. Plants with somatic mutations can be bypassed by screening for mutations in T_1 plants that lack the *SpCas9* transgene. The progeny of *SSU*-4 had a 3:1 segregation ratio as would be expected for a hemizygous insertion. All of the progeny had mutations identified in the T_0 progenitor, indicating that a bi-allelic mutation had occurred at an early stage of embryogenesis. The remaining lines had a higher rate of inheritance for the *SpCas9* transgene, suggesting that multiple insertion events may have occurred during transformation. A sufficient population of mutated transgene-free plants was not obtained in the T_1 generation in this study. However, mutated alleles will be transmitted to the subsequent generation (T_2 generation) of transgene-free plants. Therefore, isogenic transgene-free populations should be straightforward to obtain for future studies.

Targeted knockout of *rbcS-S1* and *rbcS-T1* produced plants with *ca.* 5-10% of the Rubisco present in wild-type leaves. Rubisco content can also be decreased by antisense targeting of

rbcS in tobacco (Hudson et al., 1992; Rodermel et al., 1988) and rice (Makino et al., 1997). The antisense fragment used to suppress tobacco *rbcS* was derived from *Nicotiana sylvestris* (Genbank: J01308.1) (Pinck et al., 1984; Rodermel et al., 1988). The *N. sylvestris* cDNA clone had *ca.* 97% sequence identity to the two *rbcS* clones identified in tobacco (*rbcS-S1a* and *rbcS-S1b*) (Mazur and Chui, 1985; O'Neal et al., 1987). Although the S-isoforms in tobacco originated from the *N. sylvestris* genome, the translation of *rbcS-T1* would have been suppressed owing to high nucleotide identity with *rbcS-S1*. A second group generated plants with suppressed *rbcS* transcripts using an antisense sequence derived from *N. tabacum* clone TSSU3-1 (*rbcS-S1a* and *rbcS-S1b*) (Hudson et al., 1992; Masle et al., 1993). In both cases, the suppression of *rbcS* transcripts was stably transmitted to the progeny of antisense tobacco lines (Masle et al., 1993; Quick et al., 1991a). However, the effectiveness of suppression can vary between organs and developmental stages (Stitt and Schulze, 1994). For example, the transcription of Rubisco subunits decreases during senescence in wild-type leaves (Jiang et al., 1993). However, decreased production of Rubisco does not occur in the senescent leaves of antisense plants (Miller et al., 2000). Furthermore, it can be challenging to obtain large isogenic populations for destructive experiments due to the need to characterise the extent of suppression in each plant. Therefore, the *SSU* lines with a stable decrease in Rubisco content and lacking the *SpCas9* transgene could provide a useful tool for further analysis and complementation strategies.

In the current study, small mutations in *rbcS-T1* and *rbcS-S1* did not affect the total amount of *rbcS* transcripts (i.e. in *SSU-9* and *SSU-12*). The transcript level for *rbcS-T1* could not be determined in one line (*SSU-4*) with a deletion that removed the forward primer binding site. However, in all three lines, the expression of three minor isoforms was upregulated. Elevation of *rbcS-S2*, *rbcS-S4*, and *rbcS-T5* suggests a different mode of regulation (e.g. a compensation effect) compared to the isoforms that maintained wild-type levels. A similar compensation effect was observed in antisense rice (*Oryza* sativa), where suppression of one of the major

isoforms (*OsrbcS2*) was associated with an increase in the expression of a minor isoform *OsrbcS5* (Ogawa et al., 2012). In contrast, the transcript levels of minor *rbcS* isoforms were unaffected in the Arabidopsis *1a3b* mutant (i.e. a double knockout mutant of the two major *rbcS* isoforms *RbcS1A* and *RbcS1A*) (Izumi et al., 2012). Functional compensation by specific members of other gene families has been observed in Arabidopsis (Sappl et al., 2004) and tomato (Tieman et al., 2000). Gene families arise from segmental or tandem duplication of chromosomal regions, resulting in members with different expression patterns that encode similar or identical protein products. Recently duplicated genes have accumulated less single nucleotide polymorphisms (SNPs) and are more likely to have more functional compensation than more ancient duplications (e.g. in Arabidopsis) (Hanada et al., 2009). As a relatively recent polyploid, tobacco has a smaller proportion of SNPs compared the diploid progenitors (Gong et al., 2014).

The synthesis of Rubisco subunits is tightly regulated and co-ordinated at the transcriptional and post-transcriptional level (Gutteridge and Gatenby, 1995). Constitutive expression of antisense-*rbcS* mRNA suppressed *rbcS* mRNA accumulation, with some lines reduced to *ca.* 12% of wild-type tobacco (Rodermel et al., 1996). A decreased amount of *rbcS* mRNA did not affect the accumulation of *rbcL* transcripts in tobacco or rice (Makino et al., 1997; Ogawa et al., 2012; Rodermel et al., 1996). However, the synthesis of LSU and SSU proteins was co-ordinately decreased. As the RGEN approach does not suppress *rbcS* transcripts it is expected that the mutant lines accumulate an equivalent amount of *rbcS* transcripts to wild-type. However, all lines had significantly less SSU proteins indicating lower rates of translation or degradation of mutated SSUs. Similarly, and in line with previous *RbcS* knockdown studies, the *SSU* mutant lines had less LSU protein than wild-type. A small but significant decrease in the level of *rbcL* transcripts was also observed. The transcription of *rbcL* could be affected in the *SSU* lines by differential regulatory mechanisms (Wostrikoff and Stern, 2007). Antisense lines have lower levels of *rbcS* transcripts and produce functional SSU peptides at a lower rate

of translation. However, *SSU* mutant lines *SSU-9* and *SSU-12* have an equivalent accumulation of *rbcS* transcripts to wild-type but likely produce truncated peptides. For example, expression of a *polygalacturonase* (PG) gene transcript that produces a truncated peptide in tomato reduced the accumulation of the endogenous *PG* mRNA transcript (Smith et al., 1990). Thus, it is feasible that transcription inhibition resulting from enhanced degradation of truncated SSU peptides could also affect the abundance of *rbcL* mRNA in the *SSU* lines.

Reducing Rubisco content to an average of 7% of wild-type in the leaves of *SSU* mutants severely impaired the rate of photosynthesis and biomass accumulation. A severe reduction in Rubisco content decreased the maximum photosynthetic capacity of the *SSU* lines to *ca.* 42% of wild-type levels. Reducing the Rubisco content by half in tobacco plants had a minimal effect on the maximum photosynthetic capacity under normal growth conditions (Quick et al., 1991a, 1991c; Stitt et al., 1991). Further decreases in Rubisco proportionally reduced photosynthetic rates and the timing of changes in photosynthetic capacity during leaf development (Jiang and Rodermel, 1995). Lower photosynthetic rates have been associated with increased C_i owing to higher stomatal conductance (Jiang and Rodermel, 1995; Quick et al., 1991c; von Caemmerer et al., 2004). However, antisense mutants with *ca.* 18% of wild-type Rubisco had equivalent amounts of chlorophyll per leaf area compared to wild-type (Quick et al., 1991b). In contrast, the reduced photosynthetic rates for the Rubisco SSU mutants in the current study were associated with lower amounts of leaf chlorophyll. A more severe decrease in Rubisco content (7% of wild-type levels) could also affect the accumulation of chlorophyll and other photosynthetic components. Antisense tobacco lines had a slightly higher Rubisco activation state than wild-type that could have ameliorated the impact on other photosynthetic components (Jiang and Rodermel, 1995; Quick et al., 1991c). Furthermore, reducing Rubisco content to *ca.* 20% of wild-type levels reduced chlorophyll content in Arabidopsis (Izumi et al., 2012).

To compensate for reduced photosynthetic rates, antisense lines with *ca.* 18% of wild-type Rubisco activity invested less biomass to leaf structural components (Quick et al., 1991b). Similarly, in the present study the leaf area ratio (leaf area per dry weight) was increased by two-fold in *SSU* mutants compared to wild-type. Increased investment to leaf expansion maximises the photosynthetic area to enable whole-plant growth (Quick et al., 1991a). Furthermore, less biomass allocation to non-photosynthetic tissues in favour of photosynthetic tissues would allow for a more favourable photosynthetic capacity. Decreasing Rubisco to *ca.* 60% of wild-type levels increased the shoot to root ratio but did not impair photosynthesis or leaf biomass. A proportional decrease in photosynthesis and growth occurred when Rubisco was decreased to less than 40% than in wild-type leaves.

The biomass production of the *SSU* lines was consistent with the amount of Rubisco in edited leaves. In particular, stem biomass was more severely affected than leaf biomass relative to wild-type. Similarly, plants with *ca.* 20% of wild-type Rubisco content had delayed shoot expansion (Tsai et al., 1997). Lower rates of photosynthesis prolonged the juvenile phase of shoot development by three weeks compared to wild-type. However, antisense lines attained the height of wild-type at the time of flowering. Antisense lines with less than 10% of wild-type Rubisco had a severe developmental delay and did not reach the height of wild-type plants (Jiang and Rodermel, 1995). Plant development during the flowering of *SSU* mutants was not investigated; however, the severity of Rubisco suppression could delay flowering and limit shoot development.

The targeted knockout of two *rbcS* isoforms in tobacco demonstrates the potential to engineer endogenous *rbcS* families using RGENs, such as *SpCas9*. Homozygous and bi-allelic mutants were obtained in the T_0 generation and mutated alleles were inherited in the T_1 progeny of chimeric plants. This approach may be useful for engineering *rbcS* families in species such as Arabidopsis, where isoforms are tightly linked. This study also confirms that two isoforms

account for the majority of Rubisco content in tobacco leaves. The mutants had severely reduced growth and photosynthetic rates that were consistent with results obtained from previously reported antisense lines. Furthermore, the suppression of Rubisco is consistent amongst the progeny of individual *SSU* lines. Therefore, the *SSU* mutants offer a useful isogenic population to study the Rubisco SSU, for example, by expressing non-native SSUs to complement the slow growth phenotype associated with Rubisco deficiency.

Chapter 5

Complementation of Rubisco SSU mutants with heterologous SSUs

5.1 Background

In the previous chapter, a tobacco line with reduced Rubisco content was generated that could be used as a platform for heterologous expression studies, including SSUs, LSUs and/or Rubisco assembly chaperone proteins. The Arabidopsis *1a3b* mutant (Izumi et al., 2012) has previously been used to assess the impact of non-native SSUs (Atkinson et al., 2017). But as Arabidopsis chloroplast transformation is still not well established (Ruf et al., 2019; Yu et al., 2017), *1a3b* has limited potential for testing SSUs with additional chloroplast-encoded components (e.g. LSUs and assembly chaperones). However, *1a3b* is a useful model to test and pre-select SSUs for expression in tobacco due to ease of transformation.

Expression of an heterologous SSU from the C4-plant *Sorghum bicolor* L. (sorghum) in rice produced a chimeric Rubisco enzyme with enhanced k_{cat}^{C} and only minor impairment to $S_{c/o}$ compared to wild-type rice Rubisco (Ishikawa et al., 2011). The reported k_{cat}^{C} values for Rubisco are similar between sorghum, Arabidopsis, and tobacco (Table 5-1). However, it is not known if a C4-like SSU could produce a functional chimeric enzyme in species outside of the *Poaceae* family. The Rubisco enzyme from *Limonium gibertii* L. has a higher $S_{c/o}$ compared to Arabidopsis and tobacco enzymes. Expressing Rubisco from *L. gibertii* could increase the rate of photosynthesis in tobacco depending on the compromise to k_{cat}^{C} (Galmés et al., 2005). Finally, an SSU from the *Helianthus annuus* L. (sunflower) was selected as it is similar to the Arabidopsis and tobacco enzymes. Furthermore, a chimeric enzyme composed of sunflower LSUs and tobacco SSUs can support the growth of tobacco (Kanevski et al.,

1999; Sharwood et al., 2008). Replacement of the tobacco LSU with a heterologous LSU from *H. annuus* produced a chimeric Rubisco enzyme with similar kinetic parameters to the tobacco and sunflower enzymes (Kanevski et al., 1999; Sharwood et al., 2008).

Table 5-1. Kinetic parameters of Rubisco enzymes from plants. Data shows the *in vitro* CO_2/O_2 specificity ($S_{c/o}$), maximum carboxylation rate (K_{cat}^C) and the Michaelis constant for CO_2 (K_C) determined at 25°C (Section 1.1). Data shows ± SEM where available. Abbreviations: n.d – no data, PH – cultivar Petite Havana.

Species	$S_{c/o}$ (mol mol^{-1})	K_{cat}^C (s^{-1})	K_C (μM)	**Source**
Arabidopsis thaliana	92.5±1	4.1±0.1	10.7±0.7	(Atkinson et al., 2017)
Nicotiana tabacum cv. PH	96.4 ±1.7	4.4 ± 0.2	21.2 ±2.8	(Orr et al., 2016)
Helianthus annuus	84.1±0.2	n.d	23.9±1.2	(Sharwood et al., 2008)
Limonium gibertii	112.1 ± 2.5	2.50 ± 0.1	9.1 ± 0.7	(Galmés et al., 2014)
Sorghum bicolor	70 ±1	4.09 ± 0.09	25.6±1.27	(Ishikawa et al., 2011; Jordan and Ogren, 1983)

The compatibility of three candidate SSUs was evaluated by attempting to complement the Arabidopsis *1a3b* mutant. We then developed a co-transformation strategy to remove the key native tobacco SSUs rbcS-S1 and rbcS-T1 using CRISPR/Cas and express a heterologous SSU in a single transformation event. The results highlight the use of *1a3b* as a platform to select heterologous SSUs for expression in other plant species, while the co-transformation strategy shows significant promise as a "proof-of-principle" approach to accelerate the engineering of SSUs in species that require more time-consuming transformation methods

5.2 Results

5.2.1 Heterologous SSUs can complement the Arabidopsis *1a3b* mutant

Four vectors were generated to express either the Arabidopsis *rbcS1A (1A),* the *H. annuus rbcS* (*HA*), the *L. gibertti rbcS* (*LG*) or the S. *bicolor rbcS (SB*) (Table 2-5). During assembly of the SSU expression cassettes, the mature coding sequences of *HA, LG*, and *SB* were fused to an 80 amino acid variant of the *AtrbcS1A* transit peptide, which included the 55 amino acids of the transit peptide and the first 25 amino acids of the mature *1A* sequence (Fig. 5–1). Due to time limitations we continued with transformation using these expression cassettes. However, the extended transit peptide has two advantages. The first part of the mature *AtrbcS1A* sequence enhances chloroplast localisation of heterologous proteins in tobacco chloroplasts (Kim et al., 2010). The N-terminal extension also facilitated the separation of heterologous and native SSUs by SDS-PAGE. The mature *1A* SSU is 125 amino acids in length, and has four more amino acids than the mature *LG* SSU, and two more amino acids than the mature *HA* and *SG* SSUs (Fig. 5–1B). All four *rbcS* coding sequences were driven by the *AtrbcS1A* promoter (De Almeida et al., 1989; Izumi et al., 2012). An additional set of SSU vectors were generated with C-terminal GFP fusions to confirm cellular localisation (Fig. 5–2A). GFP-tagged SSUs were transiently expressed in *N. benthamiana* leaves and fluorescence microscopy confirmed localisation of the fusion proteins to the chloroplast (Fig. 5–2B). Vectors without the GFP tag were used to transform the Arabidopsis *1a3b* mutant.

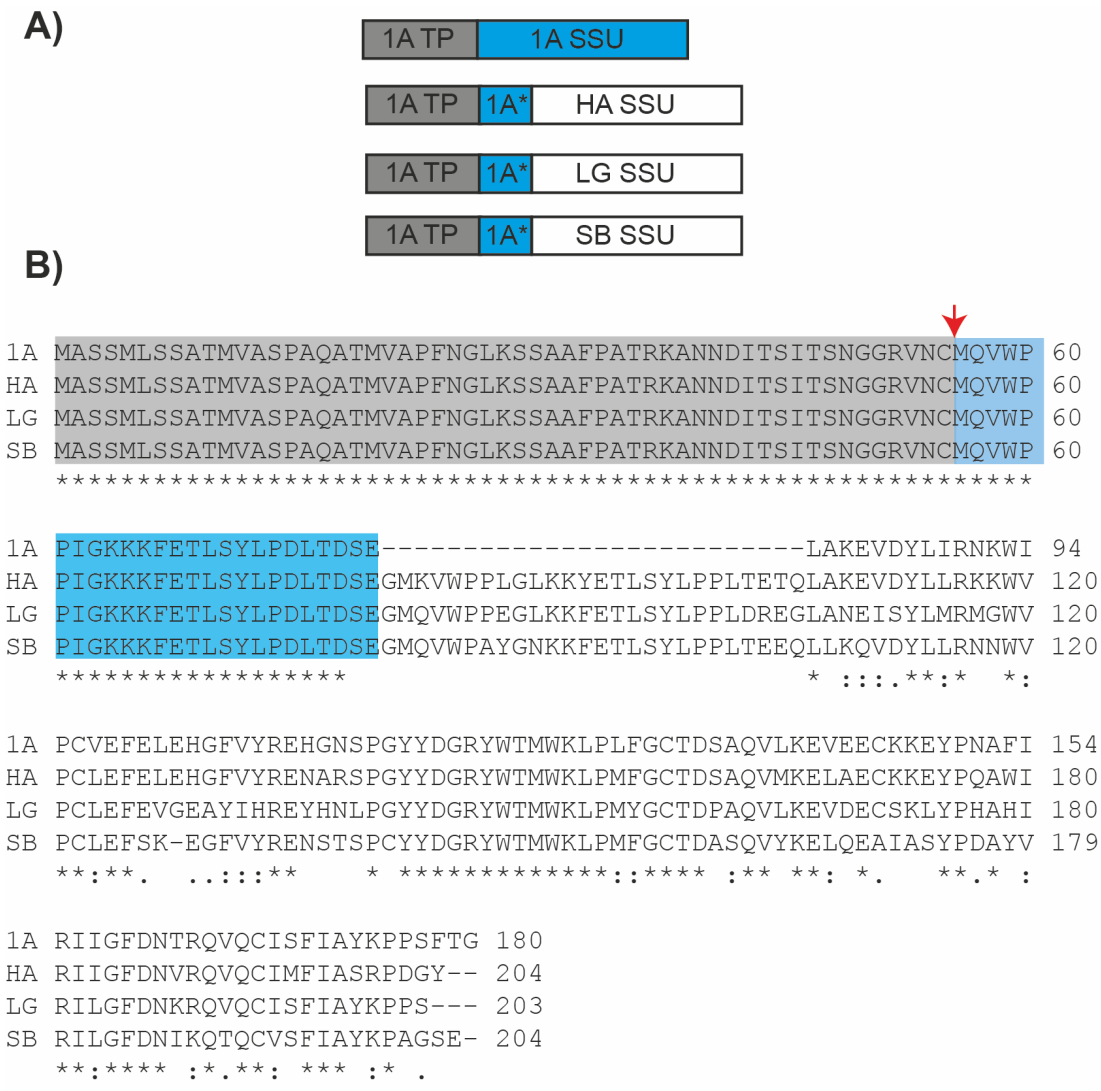


Figure 5-1. Structure of SSU peptides expressed in the Arabidopsis *1a3b* mutant. (A) The amino acid sequences of SSUs from Arabidopsis *rbcS*1A (1A), *H. annuus* (HA), *L. gibertii* (LG), and *S. bicolor* (SB) were expressed with a chloroplast transit peptide from Arabidopsis *rbcS1A* (1A TP). SB, LG, and HA were expressed with an N-terminal fusion composed of the first 25 amino acids of the mature *rbcS1A* peptide (1A*). Sources are described in Table 2-7. **(B)** Amino acid alignment of expressed proteins shows the *rbcS1A* transit peptide (grey), transit peptide cleavage site (red arrow), and N-terminal extension (blue).

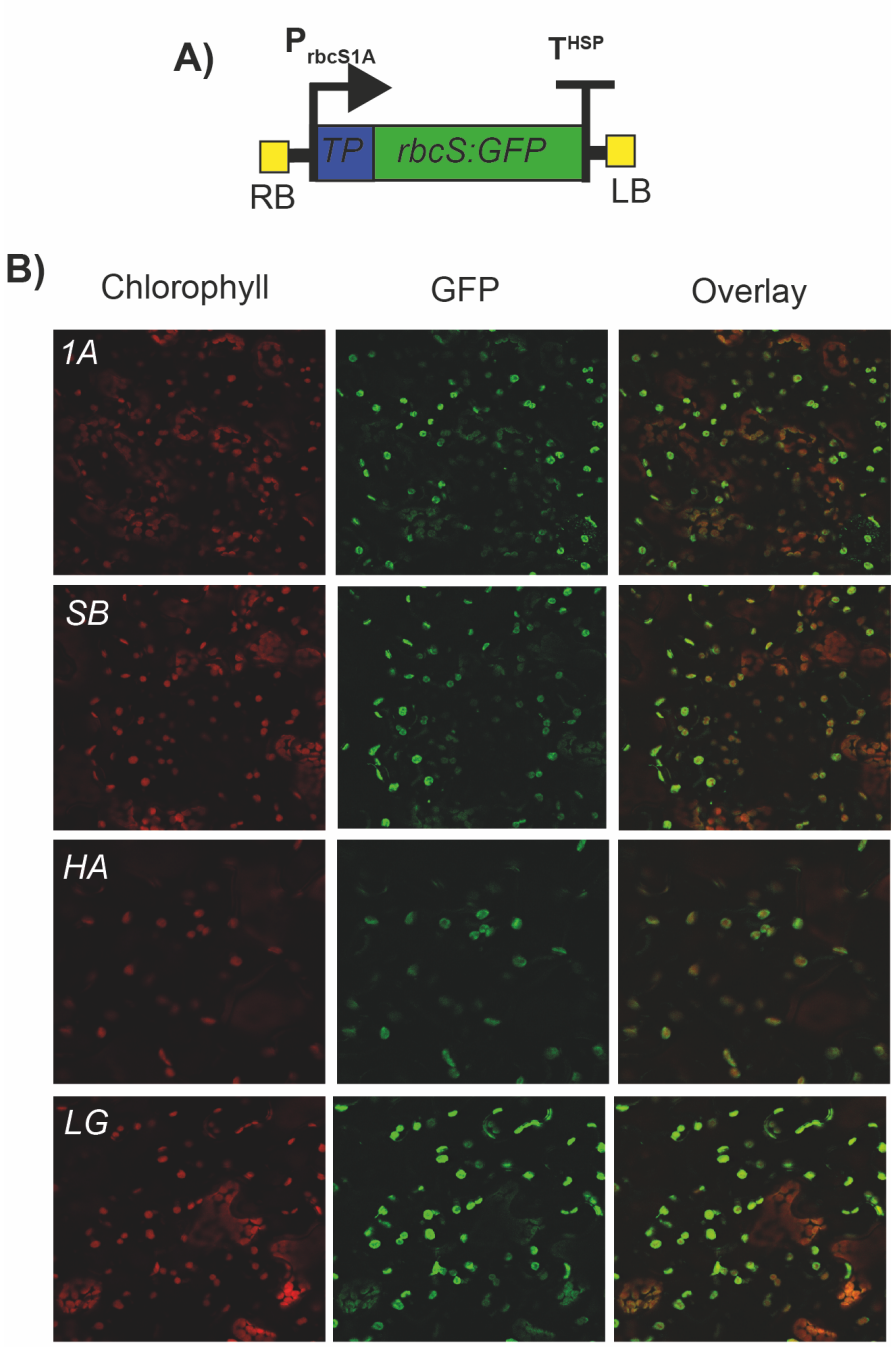


Figure 5-2. Constructs to express heterologous SSUs in Arabidopsis. (A) *rbcS* coding sequences for four SSUs (*1A, SB, HA* and *LG*) were expressed from the *AtrbcS*1A promoter (P_{rbcS1A}) and signal peptide (TP) with a GFP c-terminal tag and terminated by *AtHSP1.8* (T^{HSP}) to confirm localisation before stable expression in Arabidopsis. **(B)** Transient expression of constructs in *N. benthamiana* leaves shows GFP tagged SSUs localised to the chloroplast.

Primary transformants (T_1 generation) were identified by FAST-red seed selection (section 2.6.2). Complemented lines were selected by growing transgenic T_1 lines alongside the *1a3b* mutant. All primary *SB* and *LG* transformants had the same phenotype as the *1a3b* mutant (Fig. 5–3). In contrast, T_1 plants transformed with *1A* or *HA* were visibly larger than *1a3b*. The best performing T_1 lines for the latter were self-fertilised and T_2 seeds were collected. Homozygous and heterozygous T_2 lines were distinguished by the segregation of red to non-red (i.e., absence of the transgene) seeds. Red T_2 seeds from three separate T_1 lines were grown and self-fertilised. Three homozygous T_3 lines (i.e., all T_3 seeds were red) for the *HA* genotype (*HA-1*, *HA-2*, *HA-3*) and one homozygous T_3 line for the *1A* genotype (*1A-1*) were selected.

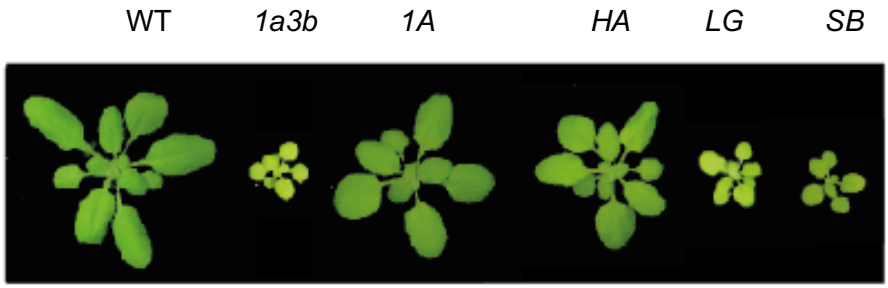


Figure 5-3. Phenotypes of Arabidopsis mutants in the T_1 generation. Wild-type (WT), *1a3b*, and *1a3b* complemented with heterologous SSUs from Arabidopsis (1A), sunflower (HA), limonium (LG) and sorghum (SG). Plants are shown 27 DAG.

Total protein was extracted from wild-type (WT), *1a3b*, *1A-1*, *HA-1*, *HA-2,* and *HA-3* (Fig. 5–4). The *1a3b* mutant had *ca.* 35% of total protein per leaf area compared to wild-type (Fig. 5–4A). Complementation with the *1A* or *HA* SSU increased the total protein to *ca.* 70% of the wild-type level with no statistically significant differences between *1A-1* and *HA-1, HA-2,* and *HA-3*. The relative amount of heterologous and native SSUs in the *HA* lines was estimated by immunoblotting (Fig. 5–4B, 4C). The intensity of bands corresponding to the heterologous (17.5 kDa) and native SSUs (14.8 kDa) was measured relative to wild-type (Fig. 5–4C). The *HA* lines had an equivalent amount of native SSUs to *1a3b* (*ca.* 30%) while the HA SSU

accounted for *ca.* 40%. No significant differences in the total amount of SSUs were detected between the *HA* lines and *1A-1*.

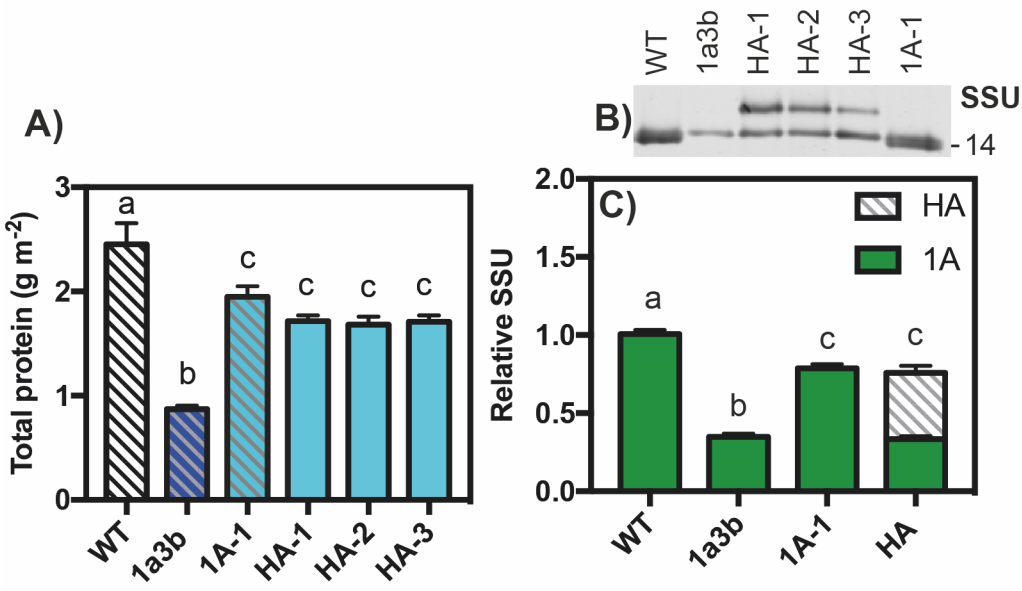


Figure 5-4. Protein analysis of Arabidopsis *1a3b* mutants complemented with a sunflower SSU (HA). (A) Total protein per leaf area determined by Bradford assay of wild-type (WT), *1a3b*, *1a3b* complemented with *rbcS1A* (1A-1) and three independent HA lines. **(B)** Immunoblotting of total protein with anti-Rubisco shows the WT Arabidopsis SSU (14.8 kDa) in all six lines and the HA SSU (17.5 kDa) in the HA lines. Equivalent amounts of protein per leaf area were loaded. Bars represent the mean ± SEM (n = 4) with different letters showing significant differences determined by ANOVA and Tukey's HSD test ($P<0.05$).

Growth experiments were performed with homozygous T_3 lines and matching *1a3b* plants from non-red seeds that were segregated from the T_2 population. The original untransformed *1a3b* mutant and WT were included as additional controls (Fig. 5–5, 5–6). As previously reported, the *1a3b* mutant had significantly lower fresh and dry weights than wild-type plants (Atkinson et al., 2017; Izumi et al., 2012). Complementation of *1a3b* with the *HA* SSU increased the dry weight to *ca.* 60% of wild-type (Fig. 5–5B). Similarly, the rosette area of the *HA* lines was *ca.* 60% of wild-type in 32-day-old plants (Fig. 5–6). The rate of rosette expansion differed between the three *HA* lines. *HA-1* was significantly smaller than wild-type

during growth. *HA-2* and *HA*-3 grew significantly slower than *HA*-1 but were not significantly smaller than wild-type until day 26 and day 29 respectively.

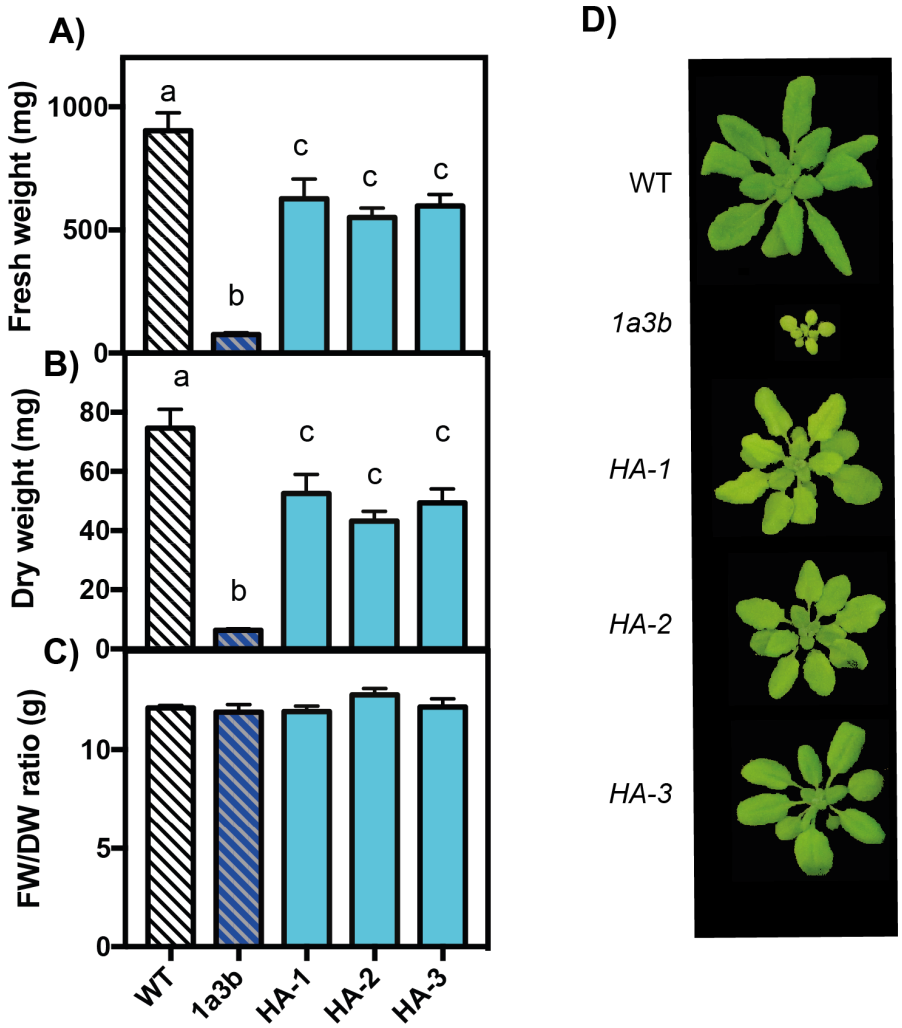
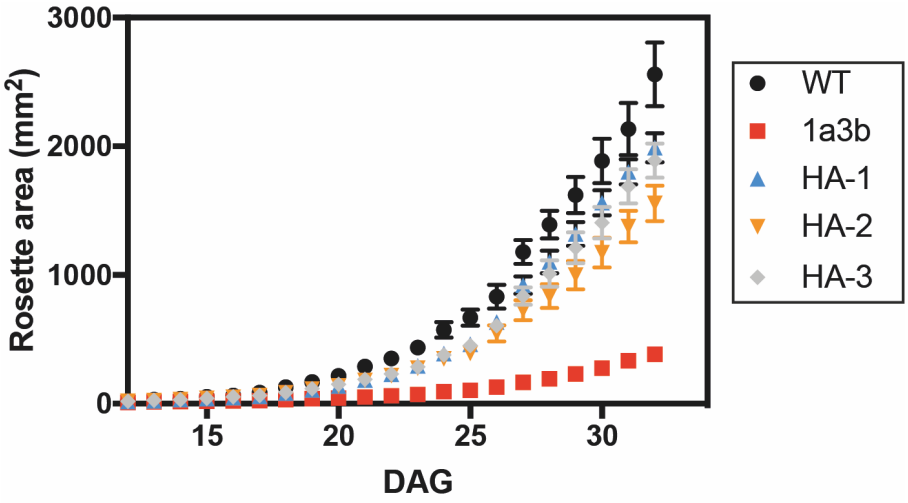


Figure 5-5. Growth of Arabidopsis *1a3b* mutants complemented with a heterologous SSU from sunflower. (A) Fresh weight of whole rosettes after 32- days of growth. **(B)** Dry weight. **(C)** Ratio of fresh weight to dry weight. Values are the mean ± SEM (n=10) followed by letters indicating significant differences determined by ANOVA followed by Tukey's HSD tests ($P<0.05$).**(D)** Phenotypes of wild-type (WT), *1a3b*, and three independent complemented lines (*HA*) after 29-days of growth under the conditions described in Fig. 5–4.



	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32
1a3b	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*
HA1	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*
HA2														*	*	*	*	*	*	*	*
HA3																	*	*	*	*	*

Figure 5-6. Rosette expansion of complemented Arabidopsis *1a3b* lines. Wild-type (WT), *1a3b* and three sunflower SSU complemented *1a3b* lines (HA) were grown in at an irradiance of 170 µmol photons m^{-2} s^{-1} in a 12-hour photoperiod. Starting 12-days after germination (DAG) plants were imaged at the same time each day for 22 days. Rosette area was determined by image analysis using IDiel Plant software (Dobrescu et al., 2017). Each data point represents the mean ±SEM (n = 15). Significant differences compared to wild-type are indicated by an asterisk as determined by an ANOVA followed by Tukey's HSD test ($P<0.05$)

The response of CO_2 assimilation (A) to changing irradiance (PAR) was measured in wild-type, one *1A* line (*1A-1*), and one *HA* line (*HA-3*) (Fig. 5–7B). The light-saturated rate of CO_2 assimilation at atmospheric CO_2 concentrations (A_{max}) was significantly higher in wild-type plants than the *1A-1* and *HA-3* lines (Table 5–2). Under saturating light, the response of A to changes in the internal CO_2 concentration (C_i) was also equivalent between wild-type and *1A-1* (Fig. 5–7A). There were no significant differences in the initial slope of the A/C_i curve, maximum rate of Rubisco carboxylation ($V_{c,max}$), the maximum rate of electron transport (J_{max})

and CO_2 compensation point (Γ) between wild-type and *1A-1* (Table 5–2). These parameters were significantly lower for *HA-3* despite equivalent amounts of total protein and SSU between *HA-1* and *1A-1*. This could potentially occur because of changes in the catalytic characteristics of Rubisco formed of *HA* SSUs and Arabidopsis LSUs.

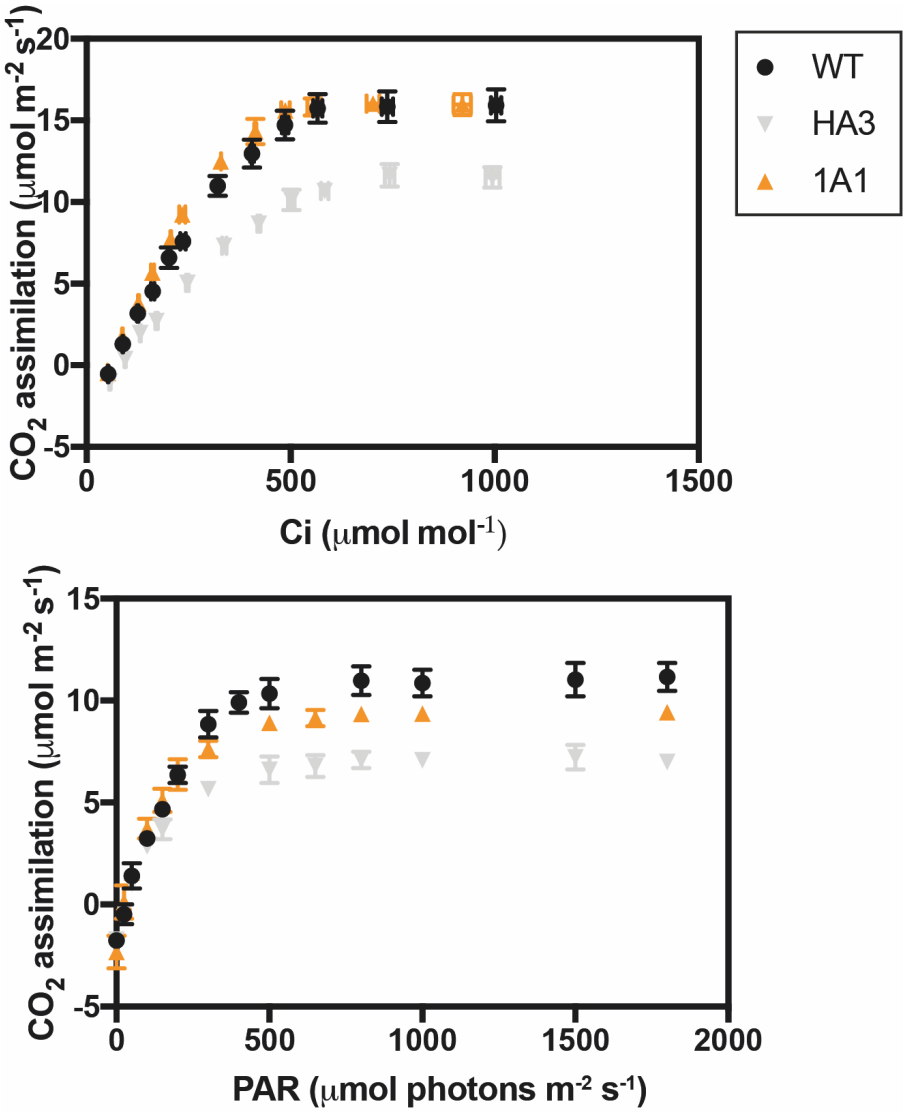


Figure 5-7. Photosynthetic response curves of complemented *1a3b* lines. (A) The response of CO_2 assimilation to intracellular CO_2 concentration (C_i) measured at 25°C under saturating irradiance (1500 µmol photons m^{-2} s^{-1}). **(B)** The response of CO_2 assimilation to changing irradiance measured at 25°C under atmospheric CO_2 concentrations (400 µmol mol^{-1}). Each data point represents the mean ±SEM of three leaves each from a separate plant.

Table 5-2. Photosynthetic parameters of Arabidopsis *1a3b* mutant complemented with heterologous SSUs. The A/Q response curves were used to derive the light-saturated rate of photosynthesis at atmospheric CO_2 (A_{max}). The maximum rate of carboxylation by Rubisco ($V_{c,max}$), maximum electron transport rate (J_{max}) and CO_2 compensation point (Γ) were estimated by fitting the A/Ci response curves. Values represent the mean ±SEM (n=3) as for Figure 7. Different letters indicate significant differences determined by ANOVA followed by Tukey's HSD tests (P<0.05).

	WT	***1A-1***	***HA-3***
A_{max} (μmol CO_2 m^{-2} s^{-1})	11.2 ± 0.5[a]	9.8 ± 0.5[b]	8.7 ± 0.4[b]
$V_{c,max}$ (μmol CO_2 m^{-2} s^{-1})	53.5 ± 0.4[a]	49.5 ±1.2[a]	31.9 ± 1.0[c]
J_{max} (mmol e^- m^{-2} s^{-1})	82.4 ± 2.5[a]	78.7 ±0.6[a]	68.6 ± 1.8[b]
Γ (μmol CO_2 mol^{-1})	57.0 ± 1.6[a]	60.1 ±1.2[a]	81.2 ± 3.4[b]
Initial slope (A/C_i)	0.052 ± 0.001[a]	0.050 ±0.001[a]	0.031 ± 0.001[c]

Successful complementation of the *1a3b* mutant with the *HA* SSU highlights the use of *1a3b* as a system to screen SSU candidates. Unfortunately, the lack of complementation by constructs with *SB* and *LG* SSUs in *1a3b* precluded their use for expression work in tobacco. Furthermore, the *HA* SSU was initially only considered a control SSU that would be of limited interest for future work on improving the performance of Rubisco. The Rubisco enzyme from *C. reinhardtii* has more divergent catalytic properties than *HA* Rubisco and has been used previously to complement *1a3b* (Atkinson et al., 2017). Therefore, one of the two *rbcS* isoforms from *C. reinhardtii* (*RBCS2B*) was selected for SSU expression in tobacco.

5.2.2 Co-transformation strategy to replace native tobacco SSUs

The tobacco *SSU* mutants described in Chapter 4 offer a useful system for heterologous SSU expression. However, owing to time limitations, transgene-free *SSU* mutants were not obtained to test complementation. An alternative, attractive strategy is simultaneous deletion of native *rbcS* genes and complementation with heterologous SSUs, which could potentially reduce the need for multiple rounds of transformation and screening. Thus, an approach was developed to replace native SSUs by co-transforming wild-type tobacco with the *SpCas9*

construct (Appendix A Fig. 11) and an optimised construct expressing the SSU encoded by *rbcS2* from *C. reinhardtii* (*CR2*).

The *AtrbcS1A* promoter is efficient for expressing heterologous SSUs in Arabidopsis *1a3b* but had not been compared to other *rbcS* promoters for expression in tobacco. In Chapter 3, the similarity between the regulatory sequences encoding *rbcS-S1* and the *rbcS-S2* promoter from *S. lycopersicum (SlrbcS2*) was described. Four vectors were generated to express a NanoLuciferase (N_{LUC}) reporter from either *AtrbcS1A*, *AtrbcS3B, SlrbcSS2* or a constitutively expressed promoter (*Nos*) (Fig. 5–8A) (Appendix A Figs. 1-4). A constitutively expressed firefly luciferase reporter (F_{LUC}) was included in as an internal control. The four vectors were transformed into tobacco mesophyll protoplasts by PEG-mediated transformation. The relative luminescence of each promoter was measured after exposing protoplasts to a low (20 µmol photons m^{-2} s^{-1}) or medium light intensity (100 µmol photons m^{-2} s^{-1}) (Fig. 5–8B, 8C). As previously reported, the *AtrbcS3B* promoter produced a significantly lower level of expression than *AtrbcS1A* (Izumi et al., 2012). Luminescence produced by the *AtrbcS1A* and *SlrbcSS2* promoters was equivalent in low light, while the *SlrbcSS2* was stronger under a medium light intensity. Based on these results, the *SlrbcS2* promoter was chosen to express *CR2* in tobacco.

A vector containing the *CR2* coding sequence fused with a C-terminal GFP tag and the 55 amino acid chloroplast transit peptide derived from *AtrbcS1A* was assembled to confirm cellular localisation (Fig. 5–8D). Tobacco leaves were agroinfiltrated with the vector and localisation to the chloroplast was confirmed by fluorescence microscopy (Fig. 5–8E). An untagged *CR2* SSU was cloned into a vector containing a hygromycin resistance cassette for stable transformation of wild-type tobacco (Fig. 5–9A) (Appendix A Fig. 10). The *SSU1* vector (Appendix A Fig. 11) was used to express *SpCas9* and gRNAs targeting *rbcS-S1* and *rbcS-T1* (Fig. 5–9B).

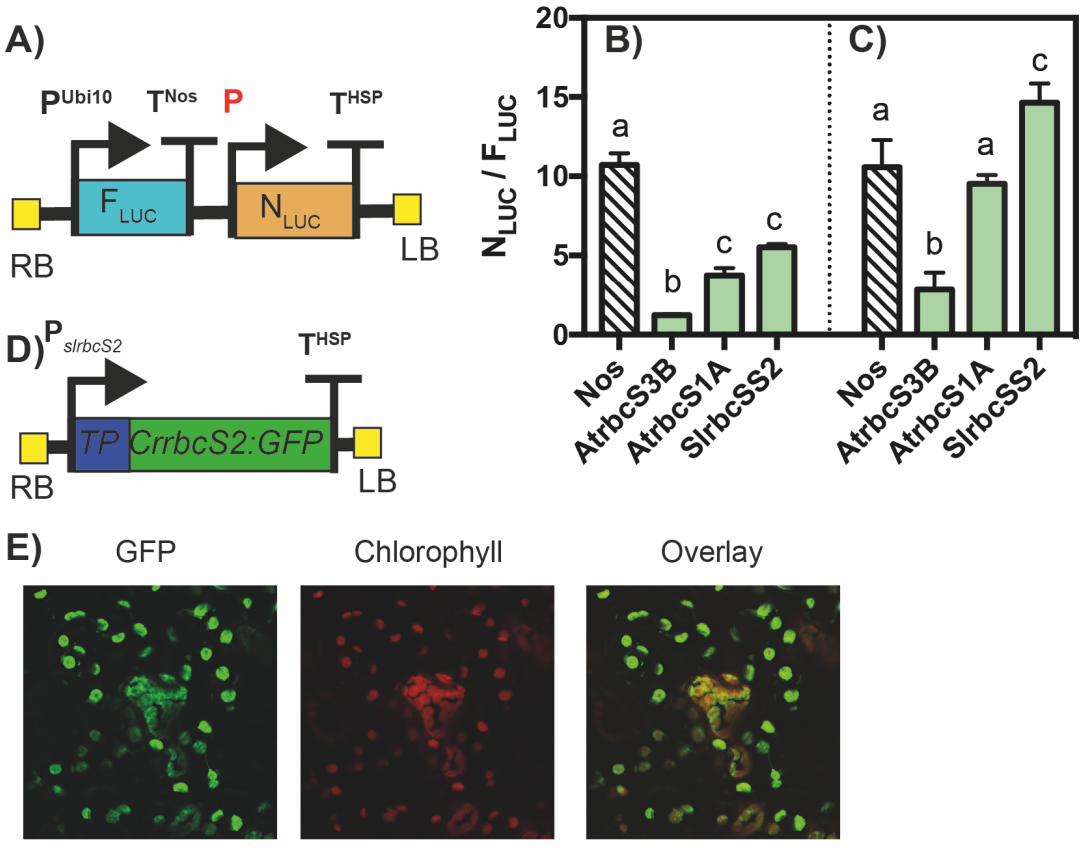


Figure 5-8. Optimisation of heterologous SSU construct for expression in tobacco. (A) Design of vectors to test promoter strength using a dual luciferase assay. Each vector contained an internal firefly luciferase reporter (F_{LUC}) control driven by a constitutive promoter (P^{Ubi10}) and a light-inducible *rbcS* promoter (P – red) driving the NanoLuc (N_{LUC}) reporter. **(B)** Vectors containing three *rbcS* promoters and one containing a control promoter (Nos) were expressed in tobacco mesophyll protoplasts. **(C)** Protoplasts were exposed to an irradiance of 20 µmol photons m^{-2} s^{-1} or 100 µmol photons m^{-2} s^{-1} of light for one hour before luciferase production was measured. The luminescence of N_{LUC} was calculated relative to the F_{LUC} internal control (i.e. N_{LUC}/F_{LUC}). Values are the mean ±SEM (n=4) followed by different letters indicating significant differences determined by ANOVA and Tukey's HSD ($P<0.05$). **(D)** The *SlrbcSS2* promoter was chosen to express the *C. reinhardtti* SSU (*CR2)* in tobacco. *CR2* SSU was transiently expressed with a GFP c-terminal tag and chloroplast signal peptide to confirm localisation. **(E)** Agroinfiltrated tobacco leaves show localisation of CR2-GFP to the chloroplast.

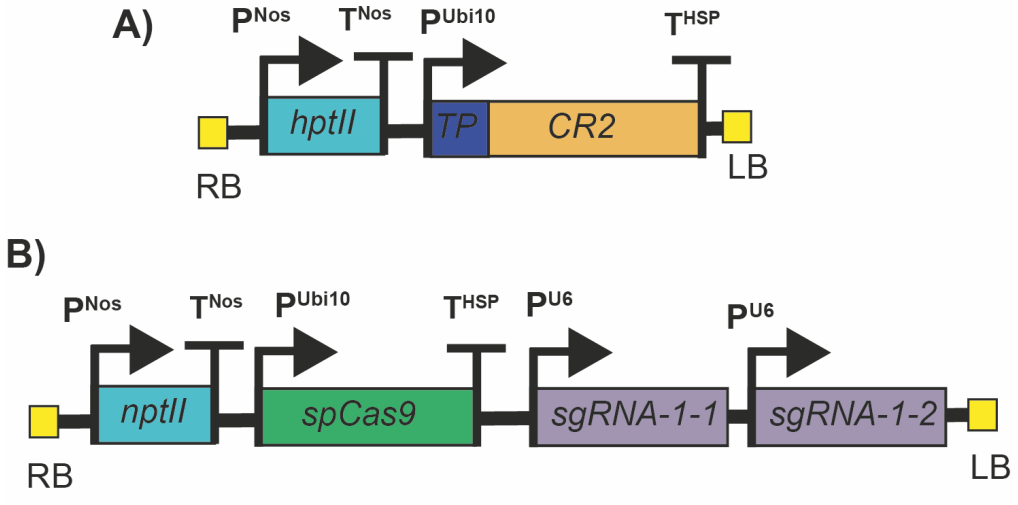


Figure 5-9. Design of constructs for co-transformation of tobacco. (A) Heterologous SSU expression construct containing the *C. reinhardtii* SSU (*CR2*) with the *AtrbcS1A* signal peptide (TP) driven by the *SlrbcS2* promoter and hygromycin selectable marker (*hptII*) to select transformed plants. **(B)** Cas9-gRNA construct to simultaneously knock-out two native *rbcS* isoforms (*rbcS-S1* and *rbcS-T1*). *SpCas9* is expressed from a constitutive promoter (P_{Ubi10}) and two gRNAs are driven by RNA polymerase III dependent promoter (P_{U6}). Transformed plants are selected by resistance to kanamycin (*nptII*).

5.2.3 Generation of tobacco plants containing the *CR2* SSU and *SpCas9* transgenes

Transgenic lines were generated by co-transformation with vectors *SSU1* and *CR2* and selected by dual resistance to kanamycin and hygromycin (Fig. 5–10). Five T_0 lines (*CT* lines) that contained both the *SpCas9* and *CR2* transgenes were identified by PCR (Fig. 5–10B). The *CT* lines had a range of phenotypes ranging from wild-type like (*CT-1* and *CT-2*) to pale leaves as observed in the *SSU* mutants (*CT-3* and *CT-4*) or a combination of both (*CT-5*) (Fig. 5–10A). As reported in Chapter 4, indel mutations are predominantly detected as opposed to a large deletion between the two gRNA in *rbcS-T1* and *rbcS-S1*. In agreement with the previous results, a large deletion was not detected in either *rbcS-S1* or *rbcS-T1* in any of the *CT* lines (Fig. 5–10C). Sequencing confirmed the presence of indels in all five plants at the second gRNA site in *rbcS-T1* (Table 5–3). Three lines, *CT-1, CT-2,* and *CT-5* were chimeric and

contained the wild-type allele in addition to mutated alleles. Multiple mutant alleles were identified in *CT-3* however, the wild-type allele was not present. One line (*CT-4*) contained two mutated alleles and so was putatively bi-allelic for *rbcS-T1*. It was not possible to sequence *rbcS-S1* amplicons in the T_0 generation (discussed in detail in Chapter 4) but mutations in *rbcS-S1* were later identified in the T_1 progeny of *CT-3* and *CT-4*.

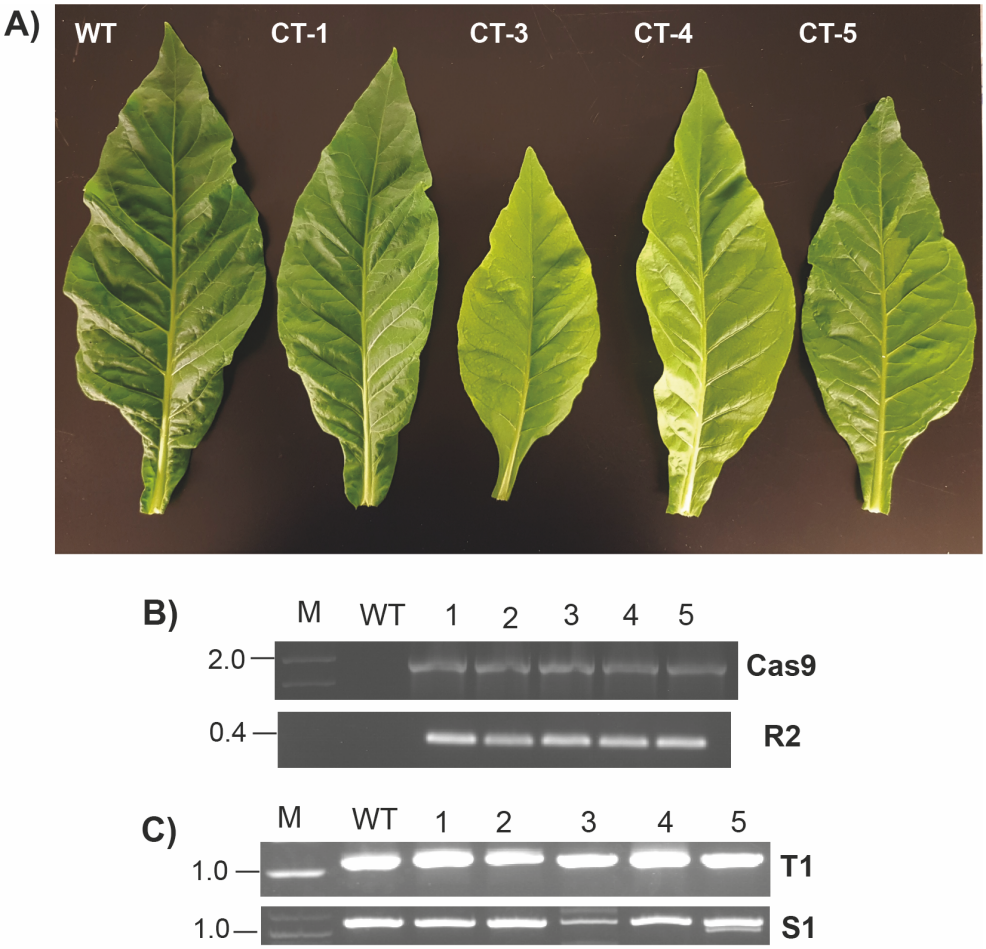


Figure 5-10. Screening co-transformed (CT) plants in the T_0 generation. (A) Phenotypes of plants co-transformed with vectors described in Fig. 5–7. Wild-type (WT) and four CT lines with phenotypes ranging from wild-type like (CT-1) to pale (CT-3 and CT-4) and chimeric (CT-5). **(B)** Confirmation of transgene integration by PCR with gene-specific primers shows *SpCas9* and *CR2* amplified from CT DNA. **(C)** Gene-specific primers amplify a 1.1 kb fragment for *rbcS-T1* and *rbcS-S1* in WT and CT lines. Amplicons from CT lines were sequenced to confirm mutations.

Table 5-3. Mutations and phenotypes of T_0 generation tobacco plants with mutations in *rbcS-T1*. Alleles were identified by Sanger sequencing of PCR products (Fig. 5–10C) and analysed using Tracking Indels by DEcomposition (TIDE) software. Indels were detected at the second gRNA site in *rbcS-T1*. WT – wild-type, d – deletion, i – insertion followed by the number of base-pairs differing from the WT allele.

Line	Phenotype	Zygosity	Alleles
CT-1	WT	Chimeric	WT, d1, i1
CT-2	WT	Chimeric	WT, d9, d1, i1
CT-3	Pale	Chimeric	d9, d2, d1, i1
CT-4	Pale	Bi-allelic*	d1, i1
CT-5	Chimeric	Chimeric	WT, d9, d1, i1

The T_1 progeny of the five *CT* lines were selected for the *CR2* transgene by hygromycin resistance (Table 5-4). All of the progeny from two lines (*CT-2* and *CT-5*) failed to germinate on hygromycin containing media. The lack of transgene inheritance could indicate instability of the transgene in the T_0 plants. The *CR2* transgene segregated at a ratio of 10:1 in the progeny of line *CT-1*, while all of the T_1 progeny of *CT-3* and *CT-4* were resistant to hygromycin. Inheritance of the *CR2* transgene was confirmed in 20 T_1 plants from *CT-3* and *CT-4*. T_1 progeny from *CT-1*, *CT-3* and *CT-4* were then screened for the *SpCas9* transgene. T_1 plants were not selected for kanamycin resistance as it is desirable to select transgene-free lines to confirm the heritability of *SpCas9* induced mutations. The *SpCas9* transgene was absent in all of the *CT-1* lines (0/20 plants), while all of the progeny of *CT-3* and *CT-4* inherited the *SpCas9* transgene (20/20 plants for each line). 15 T_1 plants from each of the three lines were screened for mutations in *rbcS-S1* and *rbcS-T1* (Table 5–3). Transgene-free *CT-1* plants were homozygous for the wild-type allele in both *rbcS-T1* and *rbcS-S1* (15/15 plants). Chimeric mutations at the second gRNA site were identified in both isoforms in all of the *CT-3* progeny (15/15 plants). Eleven *CT-4* plants had one or two mutated alleles at the second gRNA site in *rbcS-T1*. However, these lines were considered putatively homozygous or bi-allelic as all of

the progeny contained the *SpCas9* transgene. Although it was not possible to confirm the heritability of mutations in the T_1 progeny of *CT-3* and *CT-4* the wild-type allele was not detected for either isoform. Therefore, these lines were considered to represent mutant lines similar to the *SSU-12* mutant described in Chapter 4.

Table 5-4. **Inheritance of the Rubisco small subunit transgene from Chlamydomonas and CRISPR/Cas mutations in two Rubisco SSU genes (*rbcS-T1* and *rbcS-S1*) in the T_1 generation.** The inheritance of *CR2* was determined by resistance to hygromycin (n=200). Twenty seedlings from each hygromycin resistant line were transferred to soil and the genotypes of 15 plants were determined by PCR and sequencing. The number of T_1 progeny with homozygous (hom), bi-allelic (bi) and chimeric (chi) alleles is shown. Non-chimeric alleles are described as wild-type (WT), deletions (d) or insertions (i) followed by the number of base-pairs compared to wild-type and the number of progeny with a single genotype. Bi-allelic and mutated homozygous plants are putative (*) due to the presence of the *SpCas9* transgene.

Line	CR2 inheritance (%)	Phenotypes	Isoform	Zygosity	Alleles
CT-1	75 (181)	WT (20)	*rbcS-T1*	15hom	WT (15)
			rbcS-S1	15hom	WT (15)
CT-2	0	n/a	n/a	n/a	n/a
CT-3	100	Pale (20)	*rbcS-T1*	15chi	n/a
			rbcS-S1	15chi	n/a
CT-4	100	Pale (20)	*rbcS-T1*	7bi*;4hom*; 4chi	d9/i1 (7) i1 (4)
			rbcS-S1	15 chi	n/a
CT-5	0	n/a	n/a	n/a	n/a

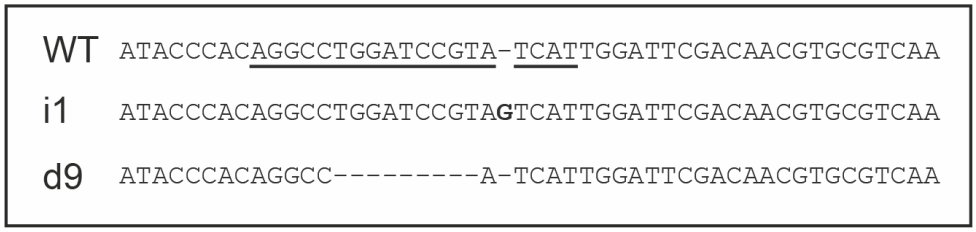


Figure 5-11. Alleles identified for *rbcS-T1* in line *CT-4*. Alleles at the second gRNA site in *rbcS-T1* identified by sequencing and TIDE analysis. 7/11 plants were putatively bi-allelic containing both a 1-bp insertion (i1) and 9-bp deletion (d9) while 4/11 plants contained the i1 allele only. The underlined region on the WT allele corresponds to the gRNA target site.

After confirming the presence of mutations in *CT-3* and *CT-4* it was necessary to establish the expression of *CR2*. First, *CR2* cDNA was detected in T_1 plants from each line by gene-specific PCR (Table 2-2) using a cDNA template (Fig. 5–12A). An amplicon corresponding to the expected size was detected in all of the transgenic plants suggesting that *CR2* is expressed. Second, total protein was extracted to confirm the expression of the heterologous SSU (Fig. 5–12B). Protein extracts from the *CR2* complemented Arabidopsis *1a3b* mutant (*R2*) were used as a positive control (Atkinson et al., 2017). The native Arabidopsis and heterologous CR2 SSUs have different mobilities on a Coomassie-stained SDS-PAGE gel, such that separation was clearly seen between CR2 (15.5 kDa) and remaining B-family SSUs RbcS1B and RbcS2B (14.8 kDa) (Fig. 5–12B). In contrast, no CR2 band was detected in extracts from *CT-3* and *CT-4* plants. Immunoblotting also failed to detect CR2 in either the *R2* or *CT* lines. Based on the presence of *CR2* cDNA it is likely that the CR2 is expressed in *CT* lines; however, at very low levels that could not be detected in this study. Optimisation of the immunoblotting procedure may be required to confirm the presence of the heterologous SSU in the *CT* mutants.

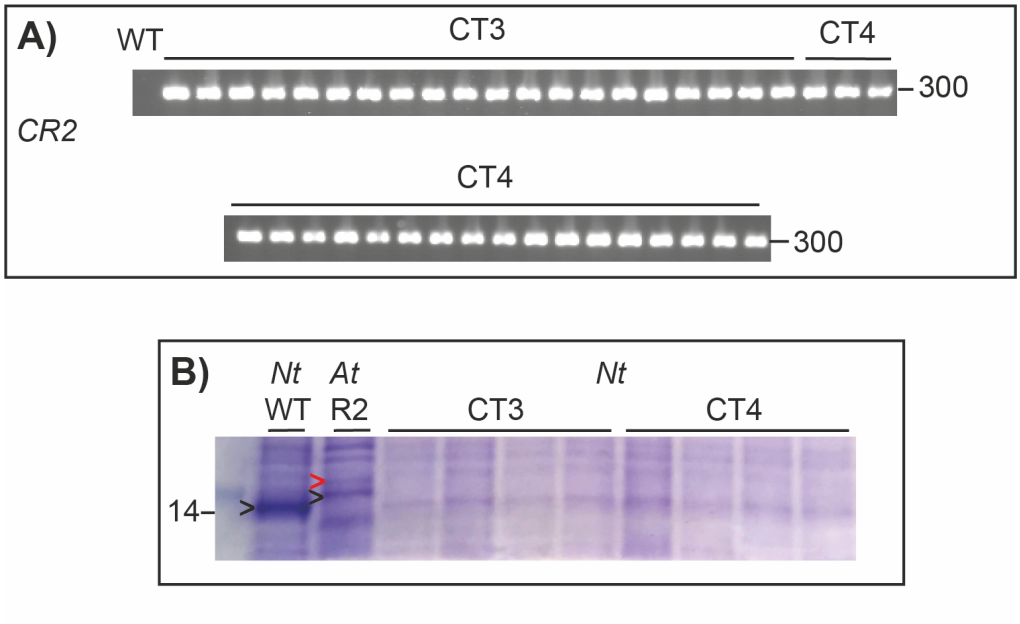


Figure 5-12. Detection of *CR2* transgene and CR2 SSU in two CT lines in the T_1 generation. (A) PCR of cDNA extracted from 20 *CT-3* and *CT-4* lines shows the presence of the *CR2* transgene in *CT* lines and not in wild-type (WT). **(B)** Coomassie stained SDS-PAGE gel of total protein shows SSU expression in WT (tobacco), the Arabidopsis *1a3b* mutant expressing *CR2* (R2) and three biological replicates of *CT-3* and *CT-4* tobacco plants. The native tobacco SSU (14.5 kDa) (black arrow) is present in WT and CT samples. The *CR2* SSU (red arrow) is detected in the R2 lines but not in *CT* plants.

Although it was not possible to detect CR2, the two lines were grown alongside wild-type to establish if growth was enhanced compared to the *SSU* mutant. Leaf discs were harvested after 45-days of growth to measure the total soluble protein (TSP) and chlorophyll content of the *CT* lines (Fig. 5–13). TSP reductions relative to wild-type were not as severe in the *CT* lines (Fig. 13C) compared to the *SSU* mutants (*ca.* 50% and 70% respectively) (Fig. 4–13). The impact on chlorophyll content was also less severe in the *CT* lines (Fig. 5–13A, 13B). Total chlorophyll was reduced by *ca.* 50% compared to wild-type (Fig. 5–13A), while the *SSU* lines had *ca.* 30% of wild-type chlorophyll per leaf area. No changes were observed in the ratio of chlorophyll *a*/*b* (Fig. 5–13B).

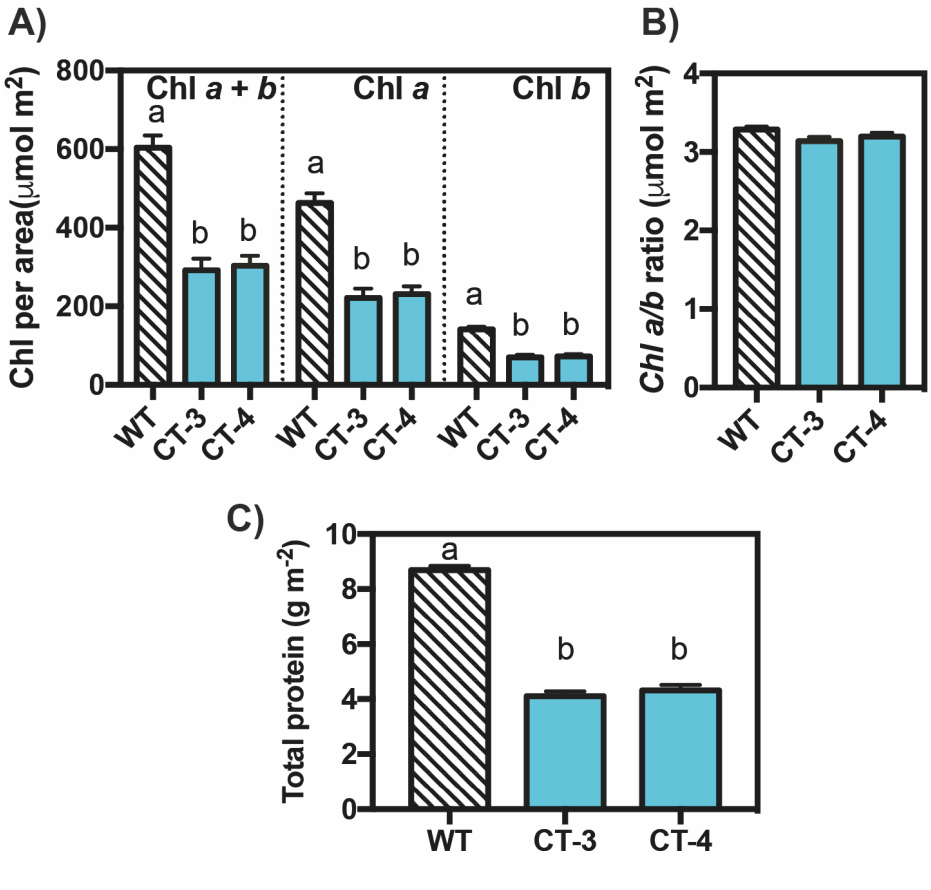


Figure 5-13. Chlorophyll and protein content in tobacco plants complemented with a heterologous SSU from *C. reinhardtti*. (**A, B**) The amount of chlorophyll *a* and chlorophyll *b* was determined in an equivalent leaf area in wild-type (WT) and two mutant lines (CT). (**C**) Total protein per leaf area determined by Bradford assay. Values represent the mean ±SEM (n=4) followed by different letters indicating significant differences determined by ANOVA and Tukey's HSD tests ($P<0.05$).

The two *CT* mutants had reduced growth rates compared to wild-type (Fig. 5–14). However, the growth impairment was less severe than for the *SSU* mutant lines. The *CT* lines were *ca.* 60% shorter than wild-type (Fig. 5–14B, 15A) and accumulated *ca.* 13% of wild-type stem biomass (Fig. 5–14C, 15B). In contrast, the *SSU* lines accumulated *ca.* 1% of stem biomass (Fig. 4–12B). The reduction in the stem dry weight accounted for more of the decrease in total dry weight than in the leaves (Fig. 5–14D, 14F). The average leaf biomass was *ca.* 40% of the wild-type and the total leaf area was reduced by *ca.* 50% (Fig. 5–14E, 15D).

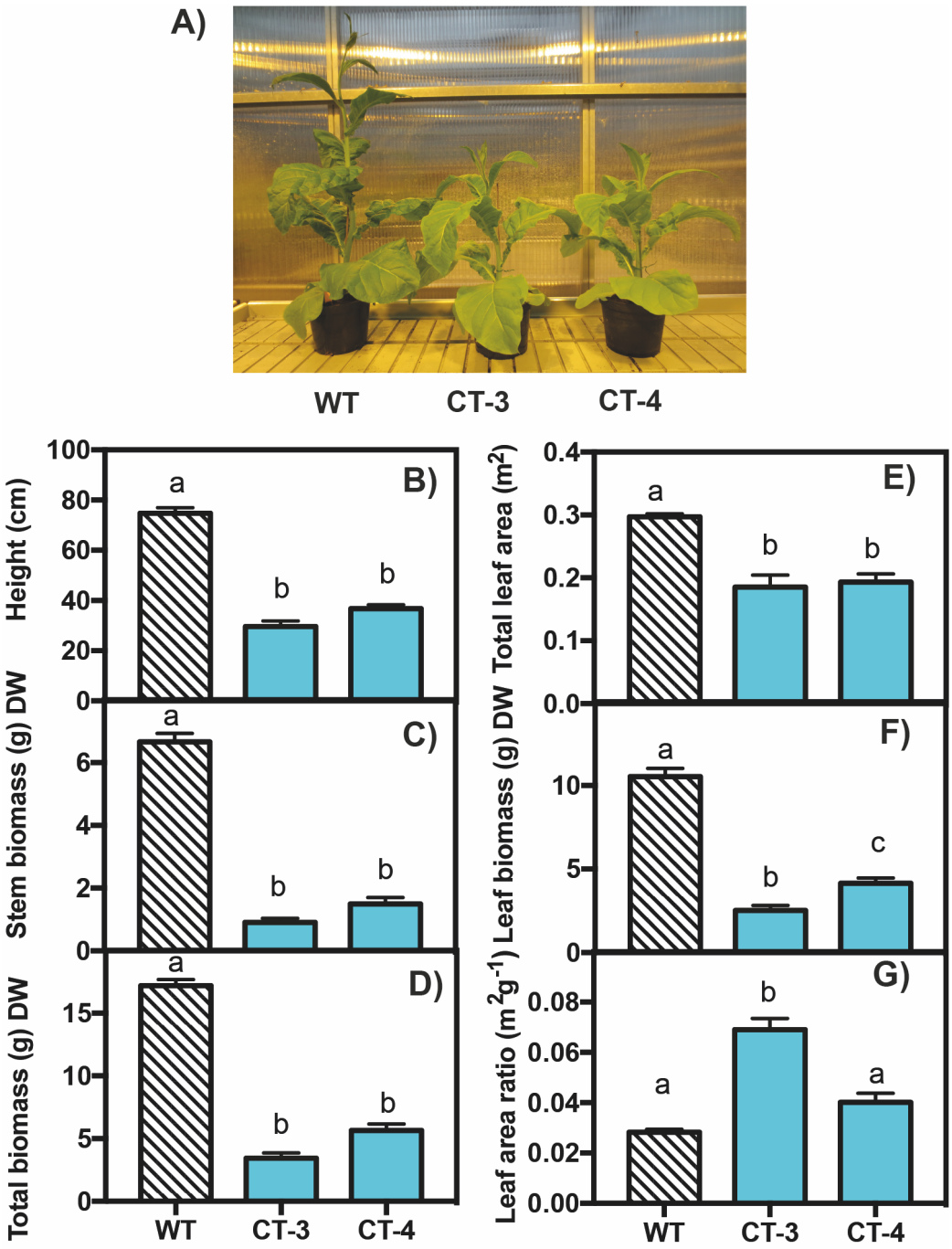


Figure 5-14. **Growth of tobacco CT mutants in the T_1 generation. (A)** Phenotypes of wild-type (WT) and two independent *CT* lines at 45-days old. **(C)** Stem dry weight. **(D)** Total dry weight. **(E)** Total leaf area. **(F)** Leaf dry weight. **(G)** Leaf area per plant dry weight. Plants were grown at 25°C in a 16 h photoperiod (200 µmol photons m^{-2} s^{-1}) for 14 days before transfer to a greenhouse with a 14 h photoperiod under natural light Values represent the mean ±SEM (n =10) followed by different letters indicating significant differences determined by ANOVA and Tukey's HSD tests ($P < 0.05$).

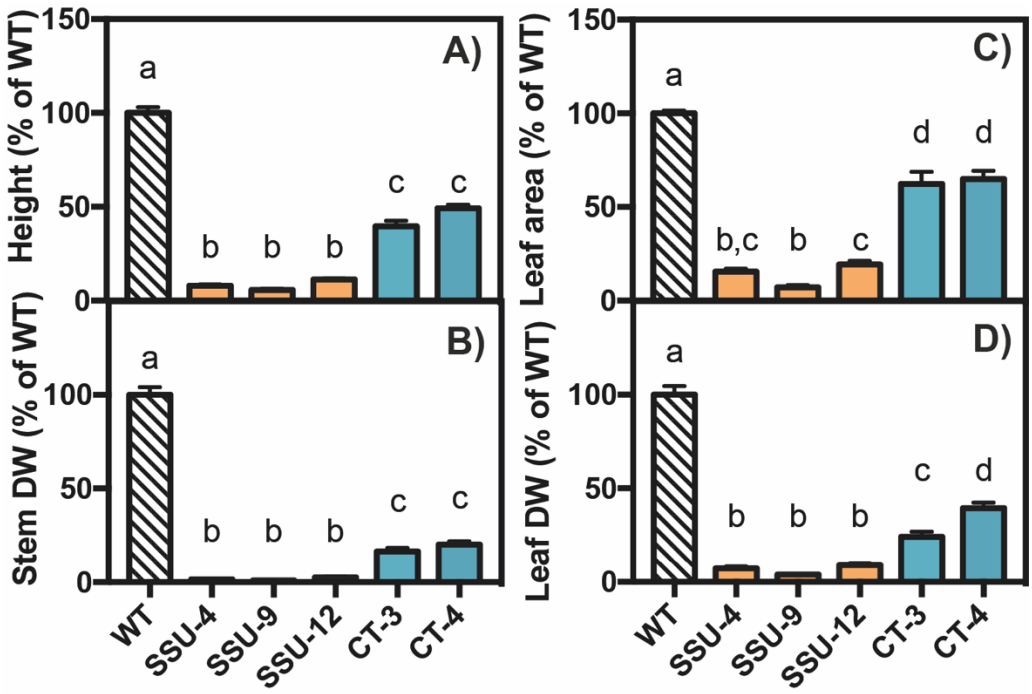


Figure 5-15. Relative growth of tobacco *SSU* and *CT* mutants. Data shows (**A)** plant height. **(B)** Stem dry weight. **(C)** Total leaf area and **(D)** Leaf dry weight from Fig. 5-14 (CT lines) and Fig. 4-12F (SSU lines) as a percentage of wild-type. Values represent the mean ±SEM followed by different letters indicating significant differences determined by ANOVA and Tukey's HSD tests ($P<0.05$).

Photosynthetic responses were measured in two independent experiments. First, A/C_i curves were measured in wild-type, *CT-3* and *CT-4* plants (Fig. 5–16). Under saturating light, the initial slope of the A/C_i response was significantly lower in *CT-3* and *CT-4* plants compared to wild-type indicating a limitation of A at low CO_2 concentrations (Table 5–5). Similarly, the parameters $V_{c,max}$ and J_{max} were significantly lower in the *CT* lines with no significant differences between *CT-3* and *CT-4.* The *A/Q* response (Fig. 5–17) was determined in the second group of plants that were grown simultaneously alongside wild-type, *SSU-4, SSU-9* and *SSU-12,* therefore, it was possible to directly compare A_{max} between the mutant lines (Fig. 5–17B). The A_{max} of the *CT* lines was *ca.* 67% of the value derived for wild-type with no significant differences between the two *CT* lines. The *SSU* mutants had a significantly lower A_{max} that corresponded to *ca.* 45% of the wild-type rate. Higher values of A_{max} in the *CT* lines

indicate a less severe limitation to *A* in atmospheric CO_2 concentrations compared to the wild-type lines.

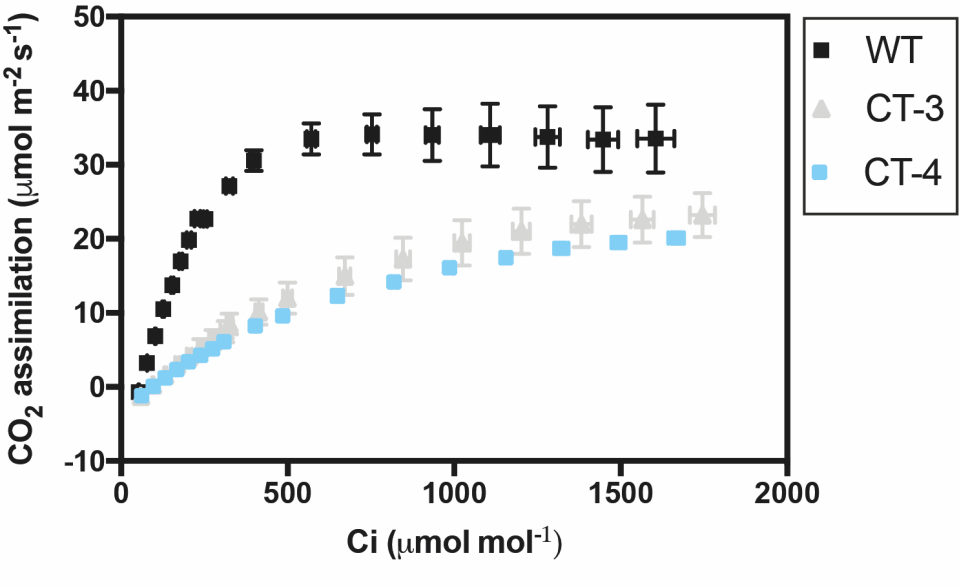


Figure 5-16. Photosynthetic response curve of co-transformed tobacco mutants. The response of CO_2 assimilation to intracellular CO_2 concentration (C_i) measured at 25°C under saturating irradiance (1500 μmol photons m^{-2} s^{-1}). Each data point represents the mean ±SEM of four leaves each from a separate plant (n=4).

Table 5-5. Photosynthetic parameters of tobacco plants expressing a heterologous SSU from *C. reinhardtii*. The maximum rate of carboxylation by Rubisco ($V_{c,max}$), maximum electron transport rate (J_{max}) and CO_2 compensation point (Γ) were estimated by fitting the *A/Ci* response curves (Fig. 10A). Values represent the mean ±SEM (n=4). Dark-adapted leaves were used for F_v/F_m measurements (n=10). Different letters indicate significant differences determined by ANOVA followed by Tukey's HSD tests (P<0.05).

	WT	***CT-3***	***CT-4***
$V_{c,max}$ (μmol CO_2 m^{-2} s^{-1})	82.9 ± 3.5^a	55.2 ± 2.7^b	56.7 ± 1.2^b
J_{max} (mmol e^- m^{-2} s^{-1})	187 ± 1.4^a	109.7 ± 6.7^b	120.2 ± 0.7^b
Γ (μmol CO_2 mol^{-1})	52.5 ± 0.3^a	89.02 ± 4.2^b	93.54 ± 1.1^b
Initial slope (A/C_i)	0.132 ± 0.003^a	0.035 ±0.008^b	0.029 ± 0.009^b
F_v/F_m	0.83 ± 0.04^a	0.80 ± 0.02^b	0.81 ± 0.01^c

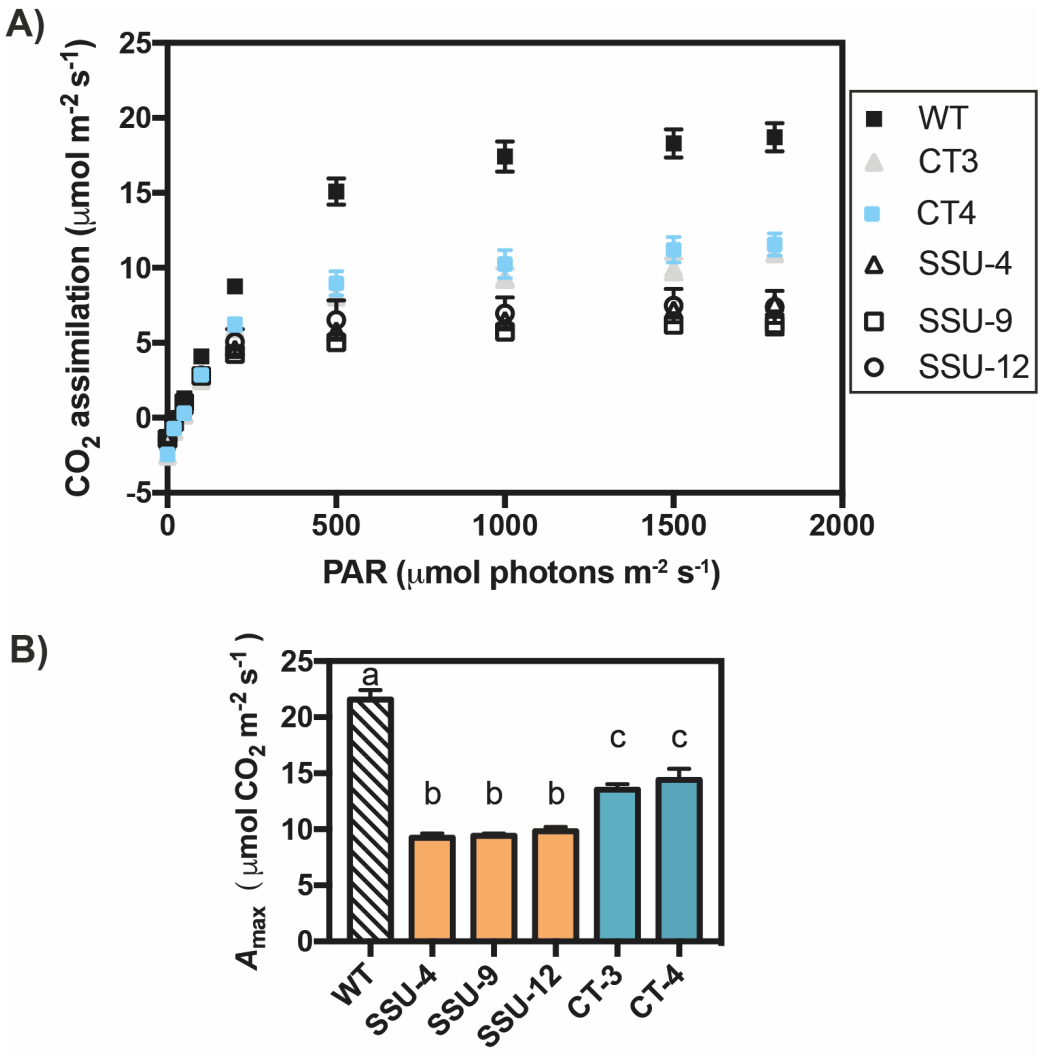


Figure 5-17. Photosynthetic responses of CT and SSU tobacco mutants. (A) The response of CO_2 assimilation (*A*) to changing irradiance (*Q*) measured at 25°C under atmospheric CO_2 concentrations (400 µmol mol^{-1}). **(B)** The *A/Q* response curves were used to derive the light-saturated rate of photosynthesis at atmospheric CO_2 (A_{max}). Each data point represents the mean ±SEM of four leaves from a separate plant (n=4). Different letters indicate significant differences determined by ANOVA followed by Tukey's HSD tests ($P<0.05$).

5.4 Discussion

This study highlights two potential approaches to examine the contribution of heterologous SSUs to the catalytic properties of Rubisco *in planta.* First, the Arabidopsis *1a3b* mutant was tested as a rapid screening tool of heterologous SSU candidates for expression in other plant species. Second, a co-transformation strategy was developed to replace the major SSUs in tobacco with a heterologous SSU in a single transformation event.

Expression of a heterologous SSU from *H. annuus* complemented the phenotype of the Rubisco *1a3b* mutant. Previously, complementation of *1a3b* with *CR2* increased the Rubisco content to *ca.* 65% of wild-type (Atkinson et al., 2017). Although Rubisco content was not measured in this study, we estimated that SSUs were restored to a similar level in the *HA* lines (*ca.* 70% of wild-type). Enhanced growth rates and photosynthetic characteristics of the *HA* lines are consistent with increased Rubisco content to at least 50% of the wild-type level (Quick et al., 1991b). This shows that the *HA* SSU can be incorporated into a functional chimeric Rubisco enzyme with Arabidopsis LSUs. The *in vitro* catalytic parameters of the chimeric Rubisco enzyme were not investigated because the k_{cat}^{c} and $S_{c/o}$ are similar between the Arabidopsis and *H. annuus* enzymes. The Rubisco enzymes from *S. bicolor* and *L. giberttii* have more interesting catalytic properties but the corresponding SSUs did not complement the *1a3b* mutant in this study. Although we did not further investigate the T_1 lines without enhanced growth rates, the lack of complementation could be because of (1) low expression in all transgenic lines, (2) lack of translation or degradation of heterologous SSUs, or (3) incompatibility between the heterologous SSUs, the Arabidopsis LSU and/or assembly chaperones. Expressing cognate assembly chaperones improves the assembly of chimeric Rubisco enzymes in *E. coli* (Aigner et al., 2017; Lin et al., 2019). However, the latter requirement is the least likely hypothesis for failure of complementation, as the *CR2* SSU, which has a lower amino acid identity to the Arabidopsis SSU than SB and LG, can be

assembled with Arabidopsis LSUs (Atkinson et al., 2017). Furthermore, the SB SSU can complement rice lines that have an antisense suppression of native SSUs (Ishikawa et al., 2011). Agroinfiltration of vectors containing GFP-tagged SSUs suggests that the coding sequences are expressed in-frame. Therefore, it is possible that lines expressing a sufficient amount of SB and LG SSUs to complement the *1a3b* mutant could be obtained by screening more primary transformants.

The data presented in this chapter shows that co-transformation can be used to disrupt two native *rbcS* isoforms in tobacco and simultaneously integrate a non-native *rbcS* isoform. This approach could be useful for lethal knock-out approaches (i.e., targeting all native SSUs) and save at least one year compared to multiple rounds of transformation. Although expression of the *CR2* SSU was not established, the *CT-3* and *CT-4* lines had a slight enhancement of growth and photosynthetic rates compared to the *SSU* mutants. Photosynthesis and growth are proportionally impaired when Rubisco content is decreased to levels of less than 40% of wild-type (Hudson et al., 1992; Quick et al., 1991b). Therefore, it could be expected that even low amounts of the CR2 SSU would proportionally improve the performance of plants lacking *rbcS-S1* and *rbcS-T1*. The *CT* mutants also had a slight increase in TSP that was consistent with increased Rubisco content (Atkinson et al., 2017). The contribution of native SSUs to the *CT* lines must also be considered. Although the wild-type allele of *rbcS-T1* and *rbcS-S1* was not detected in *CT-3* and *CT-4* most if not all of the plants had chimeric mutations. It is possible that the wild-type allele was present but not detected and could contribute to a less severe knockdown of Rubisco. The analysis of chimeric *SSU-9* and *SSU-12* mutants (Chapter 4) suggests that this is unlikely due to equivalent growth and photosynthetic rates between chimeric and homozygous or bi-allelic *SSU* mutants. In future work more T_1 lines should be examined when using this approach. However, this proof-of-principle approach demonstrates the potential for replacing entire gene families with a heterologous gene in a transformation single step in tobacco.

We have outlined an approach for testing the effect of heterologous SSUs on growth and photosynthesis in Rubisco mutants. Co-transformation of tobacco with *SpCas9* and heterologous SSU expression constructs offers an alternative strategy to replace native SSUs using a single transformation event. This represents a potentially useful method to evaluate the contribution of non-native SSUs to Rubisco following the generation of a sufficient population of primary transformants for subsequent analysis.

Final conclusions

The first aim of this work was to characterise the tobacco *rbcS* family and identify the major isoforms that contribute to Rubisco. Chapter 3 demonstrated that two isoforms, *rbcS-S1* and *rbcS-T1*, are the major isoforms expressed in tobacco. The contribution of *rbcS-S1* and *rbcS-T1* to Rubisco content was confirmed by specifically targeting these two isoforms in Chapter 4 to generate mutant lines with *ca.* 7% of wild-type Rubisco content. However, the severity of this knock-down was not anticipated based on the contributions of *rbcS-S1* and *rbcS-T1* determined in Chapter 3. A recent publication reported that *rbcS-S1a, rbcS-S1b* and *rbcS-T1* account for at least 80% of total *rbcS* transcripts (Lin *et al.,* 2019). The results of Chapter 4 confirm the contributions reported by Lin *et al.,* (2019). We predicted that both *rbcS-S1a* and *rbcS-S1b* transcripts were measured by qPCR in this study, however, amplification of a single isoform could account for our underestimation. Furthermore, both Chapter 3 and Chapter 4 highlight the challenges of distinguishing between highly homologous isoforms in tobacco. We managed to characterise eight out of at least thirteen known isoforms. Improved coverage of the tobacco will assist future efforts to distinguish between highly similar isoforms.

The second aim of this work was to generate a line with reduced expression of native SSUs. The results of Chapter 4 show that RGENs can be used to efficiently mutate two *rbcS* isoforms in tobacco. Furthermore, plants with heritable mutations (i.e. lacking the *SpCas9*) transgene can be obtained in the T_1 generation of plants. This work represents the first example of targeting the *rbcS* family using RGENs, which could be applied to other species in future work. The three *SSU* mutant lines had equivalent decreases in Rubisco content, growth and photosynthesis. A homogenous population of plants with decreased Rubisco content is particularly useful for future efforts to express introduce heterologous SSUs. We expect that

the *SSU* lines will provide a useful platform (similar to *1a3b*) for future work that involves both engineering both nuclear- and chloroplast- encoded Rubisco.

Finally, this study aimed to examine the impact of heterologous SSUs on plant growth and photosynthesis of Rubisco mutant lines. It was not possible to transform the *SSU* mutant with heterologous SSUs due to time limitations. However, Chapter 5 demonstrates a proof-of-principle approach that could be applied to examine different heterologous SSUs and/or generate a more severe knockout of native SSUs. The co-transformation approach has a clear advantage in terms of generating a complemented line in a single transformation event. A more interesting use of this system could be generating a knock-out of all *rbcS* isoforms, which would usually be lethal. Simultaneous complementation with a heterologous SSU might result in a non-lethal knockout producing a plant expressing heterologous SSUs only. The strategies outlined in Chapter 5 could contribute to efforts to express non-native Rubisco enzymes in plants.

References

Aigner, H., Wilson, R.H., Bracher, A., Calisse, L., Bhat, J.Y., Hartl, F.U., Hayer-Hartl, M., 2017. Plant RuBisCo assembly in E. coli with five chloroplast chaperones including BSD2. Science 358, 1272–1278.

Alkan, C., Sajjadian, S., Eichler, E.E., 2011. Limitations of next-generation genome sequence assembly. Nature Methods 8, 61–65.

Alonso, H., Blayney, M.J., Beck, J.L., Whitney, S.M., 2009. Substrate-induced Assembly of Methanococcoides burtonii d-Ribulose-1,5-bisphosphate Carboxylase/Oxygenase Dimers into Decamers. J. Biol. Chem. 284, 33876–33882.

Andersson, I., 2008. Catalysis and regulation in Rubisco. J Exp Bot 59, 1555–1568.

Andersson, I., Backlund, A., 2008. Structure and function of Rubisco. Plant Physiology and Biochemistry, Plant structural biology 46, 275–291.

Andersson, I., Knight, S., Schneider, G., Lindqvist, Y., Lundqvist, T., Brändén, C.-I., Lorimer, G.H., 1989. Crystal structure of the active site of ribulose-bisphosphate carboxylase. Nature 337, 229–234.

Atkinson, N., Leitão, N., Orr, D.J., Meyer, M.T., Carmo-Silva, E., Griffiths, H., Smith, A.M., McCormick, A.J., 2017. Rubisco small subunits from the unicellular green alga Chlamydomonas complement Rubisco-deficient mutants of Arabidopsis. New Phytol. 214, 655–667.

Bae, S., Park, J., Kim, J.-S., 2014. Cas-OFFinder: a fast and versatile algorithm that searches for potential off-target sites of Cas9 RNA-guided endonucleases. Bioinformatics 30, 1473–1475.

Baker, N.R., 2008. Chlorophyll Fluorescence: A Probe of Photosynthesis In Vivo. Annual Review of Plant Biology 59, 89–113.

Bar-On, Y.M., Milo, R., 2019. The global mass and average rate of rubisco. PNAS 116, 4738–4743.

Berry, J.O., Nikolau, B.J., Carr, J.P., Klessig, D.F., 1985. Transcriptional and post-transcriptional regulation of ribulose 1,5-bisphosphate carboxylase gene expression in light- and dark-grown amaranth cotyledons. Molecular and Cellular Biology 5, 2238–2246.

Berry-Lowe, S.L., Meagher, R.B., 1985. Transcriptional regulation of a gene encoding the small subunit of ribulose-1,5-bisphosphate carboxylase in soybean tissue is linked to the phytochrome response. Mol Cell Biol 5, 1910–1917.

Bracher, A., Starling-Windhof, A., Hartl, F.U., Hayer-Hartl, M., 2011. Crystal structure of a chaperone-bound assembly intermediate of form I Rubisco. Nat. Struct. Mol. Biol. 18, 875–880.

Briggs, W.R., Beck, C.F., Cashmore, A.R., Christie, J.M., Hughes, J., Jarillo, J.A., Kagawa, T., Kanegae, H., Liscum, E., Nagatani, A., Okada, K., Salomon, M., Rüdiger, W., Sakai, T., Takano, M., Wada, M., Watson, J.C., 2001. The Phototropin Family of Photoreceptors. The Plant Cell 13, 993–997.

Brinkman, E.K., Chen, T., Amendola, M., van Steensel, B., 2014. Easy quantitative assessment of genome editing by sequence trace decomposition. Nucleic Acids Res 42, e168.

Brooks, C., Nekrasov, V., Lippman, Z.B., Eck, J.V., 2014. Efficient Gene Editing in Tomato in the First Generation Using the Clustered Regularly Interspaced Short Palindromic Repeats/CRISPR-Associated9 System. Plant Physiology 166, 1292–1297.

Cashmore, A.R., Jarillo, J.A., Wu, Y.-J., Liu, D., 1999. Cryptochromes: Blue Light Receptors for Plants and Animals. Science 284, 760–765.

Cheng, S.-H., Moore, B. d, Seemann, J.R., 1998. Effects of Short- and Long-Term Elevated CO2 on the Expression of Ribulose-1,5-Bisphosphate Carboxylase/Oxygenase Genes and Carbohydrate Accumulation in Leaves of Arabidopsis thaliana (L.) Heynh. Plant Physiology 116, 715–723.

Chupeau, M.-C., Granier, F., Pichon, O., Renou, J.-P., Gaudin, V., Chupeau, Y., 2013. Characterization of the Early Events Leading to Totipotency in an Arabidopsis Protoplast Liquid Culture by Temporal Transcript Profiling. The Plant Cell 25, 2444–2463.

Clarkson, J.J., Lim, K.Y., Kovarik, A., Chase, M.W., Knapp, S., Leitch, A.R., 2005. Long-term genome diploidization in allopolyploid Nicotiana section Repandae (Solanaceae). New Phytologist 168, 241–252.

Cong, L., Ran, F.A., Cox, D., Lin, S., Barretto, R., Habib, N., Hsu, P.D., Wu, X., Jiang, W., Marraffini, L.A., Zhang, F., 2013. Multiplex Genome Engineering Using CRISPR/Cas Systems. Science 339, 819–823.

De Almeida, E.R.P., Gossele, V., Muller, C.G., Dockx, J., Reynaerts, A., Botterman, J., Krebbers, E., Timko, M.P., 1989. Transgenic expression of two marker genes under the control of an Arabidopsis rbcS promoter: Sequences encoding the Rubisco transit peptide increase expression levels. Mol Gen Genet 218, 78–86.

Dean, C., Elzen, P. van den, Tamaki, S., Dunsmuir, P., Bedbrook, J., 1985. Differential expression of the eight genes of the petunia ribulose bisphosphate carboxylase small subunit multi-gene family. EMBO J 4, 3055–3061.

Dean, C., Pichersky, E., Dunsmuir, P., 1989. Structure, Evolution, and Regulation of RbcS Genes in Higher Plants. Annual Review of Plant Physiology and Plant Molecular Biology 40, 415–439.

Dean, C., van den Elzen, P., Tamaki, S., Black, M., Dunsmuir, P., Bedbrook, J., 1987. Molecular characterization of the rbcS multi-gene family of Petunia (Mitchell). Mol Gen Genet 206, 465–474.

Dedonder, A., Rethy, R., Fredericq, H., Montagu, M.V., Krebbers, E., 1993. Arabidopsis rbcS Genes Are Differentially Regulated by Light. Plant Physiology 101, 801–808.

Denton, J.F., Lugo-Martinez, J., Tucker, A.E., Schrider, D.R., Warren, W.C., Hahn, M.W., 2014. Extensive Error in the Number of Genes Inferred from Draft Genome Assemblies. PLOS Computational Biology 10, e1003998.

Dhingra, A., Portis, A.R., Daniell, H., 2004. Enhanced translation of a chloroplast-expressed RbcS gene restores small subunit levels and photosynthesis in nuclear RbcS antisense plants. PNAS 101, 6315–6320.

Dobrescu, A., Scorza, L.C.T., Tsaftaris, S.A., McCormick, A.J., 2017. A "Do-It-Yourself" phenotyping system: measuring growth and morphology throughout the diel cycle in rosette shaped plants. Plant Methods 13, 95.

Doll, N.M., Gilles, L.M., Gérentes, M.-F., Richard, C., Just, J., Fierlej, Y., Borrelli, V.M.G., Gendrot, G., Ingram, G.C., Rogowsky, P.M., Widiez, T., 2019. Single and multiple gene knockouts by CRISPR–Cas9 in maize. Plant Cell Rep 38, 487–501.

Durr, J., Papareddy, R., Nakajima, K., Gutierrez-Marcos, J., 2018. Highly efficient heritable targeted deletions of gene clusters and non-coding regulatory regions in Arabidopsis using CRISPR/Cas9. Scientific Reports 8, 4443.

Edwards, K.D., Fernandez-Pozo, N., Drake-Stowe, K., Humphry, M., Evans, A.D., Bombarely, A., Allen, F., Hurst, R., White, B., Kernodle, S.P., Bromley, J.R., Sanchez-Tamburrino, J.P., Lewis, R.S., Mueller, L.A., 2017. A reference genome for Nicotiana tabacum enables map-based cloning of homeologous loci implicated in nitrogen utilization efficiency. BMC Genomics 18, 448.

Ellis, R.J., 1979. The most abundant protein in the world. Trends in Biochemical Sciences 4, 241–244.

Endo, M., Mikami, M., Toki, S., 2015. Multigene Knockout Utilizing Off-Target Mutations of the CRISPR/Cas9 System in Rice. Plant Cell Physiol 56, 41–47.

Engler, C., Youles, M., Gruetzner, R., Ehnert, T.-M., Werner, S., Jones, J.D.G., Patron, N.J., Marillonnet, S., 2014. A Golden Gate Modular Cloning Toolbox for Plants. ACS Synth. Biol. 3, 839–843.

Ethier, G.J., Livingston, N.J., 2004. On the need to incorporate sensitivity to CO2 transfer conductance into the Farquhar–von Caemmerer–Berry leaf photosynthesis model. Plant, Cell & Environment 27, 137–153.

Farquhar, G.D., von Caemmerer, S., Berry, J.A., 1980. A biochemical model of photosynthetic CO2 assimilation in leaves of C3 species. Planta 149, 78–90.

Feng, Z., Mao, Y., Xu, N., Zhang, B., Wei, P., Yang, D.-L., Wang, Z., Zhang, Z., Zheng, R., Yang, L., Zeng, L., Liu, X., Zhu, J.-K., 2014. Multigeneration analysis reveals the inheritance, specificity, and patterns of CRISPR/Cas-induced gene modifications in Arabidopsis. PNAS 111, 4632–4637.

Fichtner, K., Quick, W.P., Schulze, E.-D., Mooney, H.A., Rodermel, S.R., Bogorad, L., Stitt, M., 1993. Decreased ribulose-1,5-bisphosphate carboxylase-oxygenase in transgenic tobacco transformed with “antisense” rbcS. Planta 190, 1–9.

Fitchen, J.H., Knight, S., Andersson, I., Branden, C.-I., McIntosh, L., 1990. Residues in Three Conserved Regions of the Small Subunit of Ribulose-1,5- Bisphosphate Carboxylase/Oxygenase are Required for Quaternary Structure. Proceedings of the National Academy of Sciences of the United States of America 87, 5768–5772.

Fluhr, R., Chua, N.-H., 1986. Developmental regulation of two genes encoding ribulose-bisphosphate carboxylase small subunit in pea and transgenic petunia plants: Phytochrome response and blue-light induction. Proc Natl Acad Sci U S A 83, 2358–2362.

Fritz, C.C., Herget, T., Wolter, F.P., Schell, J., Schreier, P.H., 1991. Reduced steady-state levels of rbcS mRNA in plants kept in the dark are due to differential degradation. Proc Natl Acad Sci U S A 88, 4458–4462.

Fritz, C.C., Wolter, F.P., Schenkemeyer, V., Herget, T., Schreier, P.H., 1993. The gene family encoding the ribulose-(1,5)-bisphosphate carboxylase/oxygenase (Rubisco) small subunit of potato. Gene 137, 271–274.

Furbank, R., Taylor, W., 1995. Regulation of Photosynthesis in C3 and C4 Plants: A Molecular Approach. Plant Cell 7, 797–807.

Galmés, J., Andralojc, P.J., Kapralov, M.V., Flexas, J., Keys, A.J., Molins, A., Parry, M.A.J., Conesa, M.À., 2014. Environmentally driven evolution of Rubisco and improved photosynthesis and growth within the C3 genus Limonium (Plumbaginaceae). New Phytologist 203, 989–999.

Galmés, J., Flexas, J., Keys, A.J., Cifre, J., Mitchell, R. a. C., Madgwick, P.J., Haslam, R.P., Medrano, H., Parry, M. a. J., 2005. Rubisco specificity factor tends to be larger in plant species from drier habitats and in species with persistent leaves. Plant, Cell & Environment 28, 571–579.

Gao, J., Wang, G., Ma, S., Xie, X., Wu, X., Zhang, X., Wu, Y., Zhao, P., Xia, Q., 2015. CRISPR/Cas9-mediated targeted mutagenesis in Nicotiana tabacum. Plant Mol Biol 87, 99–110.

Gatenby, A.A., Vies, S.M. van der, Bradley, D., 1985. Assembly in E. coli of a functional multi-subunit ribulose bisphosphate carboxylase from a blue-green alga. Nature 314, 617–620.

Genkov, T., Meyer, M., Griffiths, H., Spreitzer, R.J., 2010. Functional Hybrid Rubisco Enzymes with Plant Small Subunits and Algal Large Subunits J. Biol. Chem. 285, 19833–19841. https://doi.org/10.1074/jbc.M110.124230

Getzoff, T.P., Bohnert, H.J., Jensen, R.G., 1998. Chimeric Arabidopsis thaliana ribulose-1,5-bisphosphate carboxylase/oxygenase containing a pea small subunit protein is compromised in carbamylation. Plant Physiology 116, 695-702.

Gong, L., Olson, M., Wendel, J.F., 2014. Cytonuclear Evolution of Rubisco in Four Allopolyploid Lineages. Mol Biol Evol 31, 2624–2636.

Gong, L., Salmon, A., Yoo, M.-J., Grupp, K.K., Wang, Z., Paterson, A.H., Wendel, J.F., 2012. The Cytonuclear Dimension of Allopolyploid Evolution: An Example from Cotton Using Rubisco. Mol Biol Evol 29, 3023–3036.

Grefen, C., Donald, N., Hashimoto, K., Kudla, J., Schumacher, K., Blatt, M.R., 2010. A ubiquitin-10 promoter-based vector set for fluorescent protein tagging facilitates temporal stability and native protein distribution in transient and stable expression studies. The Plant Journal 64, 355–365.

Gutteridge, S., Gatenby, A., 1995. Rubisco Synthesis, Assembly, Mechanism, and Regulation. Plant Cell 7, 809–819.

Gutteridge, S., Sigal, I., Thomas, B., Arentzen, R., Cordova, A., Lorimer, G., 1984. A site-specific mutation within the active site of ribulose-1,5-bisphosphate carboxylase of Rhodospirillum rubrum. EMBO J 3, 2737–2743.

Hanada, K., Kuromori, T., Myouga, F., Toyoda, T., Li, W.-H., Shinozaki, K., 2009. Evolutionary Persistence of Functional Compensation by Duplicate Genes in Arabidopsis. Genome Biol Evol 1, 409–414.

Hanada, K., Zou, C., Lehti-Shiu, M.D., Shinozaki, K., Shiu, S.-H., 2008. Importance of Lineage-Specific Expansion of Plant Tandem Duplicates in the Adaptive Response to Environmental Stimuli. Plant Physiology 148, 993–1003.

Hanson, D.T., 2016. Breaking the rules of Rubisco catalysis. J Exp Bot 67, 3180–3182.

Harpel, M.R., Hartman, F.C., 1994. Chemical Rescue by Exogenous Amines of a Site-Directed Mutant of Ribulose 1,5-Bisphosphate Carboxylase/Oxygenase That Lacks a Key Lysyl Residue. Biochemistry 33, 5553–5561.

Hauser, T., Bhat, J.Y., Miličić, G., Wendler, P., Hartl, F.U., Bracher, A., Hayer-Hartl, M., 2015a. Structure and mechanism of the Rubisco-assembly chaperone Raf1. Nature Structural & Molecular Biology 22, 720–728.

Hauser, T., Popilka, L., Hartl, F.U., Hayer-Hartl, M., 2015b. Role of auxiliary proteins in Rubisco biogenesis and function. Nature Plants 1, 15065.

Horsch, R.B., Fry, J., Hoffmann, N., Neidermeyer, J., Rogers, S.G., Fraley, R.T., 1989. Leaf disc transformation, in: Gelvin, S.B., Schilperoort, R.A., Verma, D.P.S. (Eds.), Plant Molecular Biology Manual. Springer Netherlands, Dordrecht, pp. 63–71.

Howe, C.J., Auffret, A.D., Doherty, A., Bowman, C.M., Dyer, T.A., Gray, J.C., 1982. Location and nucleotide sequence of the gene for the proton-translocating subunit of wheat chloroplast ATP synthase. Proc Natl Acad Sci U S A 79, 6903–6907.

Hudson, G.S., Evans, J.R., Caemmerer, S. von, Arvidsson, Y.B.C., Andrews, T.J., 1992. Reduction of Ribulose-1,5-Bisphosphate Carboxylase/Oxygenase Content by Antisense RNA Reduces Photosynthesis in Transgenic Tobacco Plants. Plant Physiology 98, 294–302.

Ishikawa, C., Hatanaka, T., Misoo, S., Fukayama, H., 2009. Screening of High k_{cat} Rubisco among Poaceae for Improvement of Photosynthetic CO_2 Assimilation in Rice. Plant Production Science 12, 345–350.

Ishikawa, C., Hatanaka, T., Misoo, S., Miyake, C., Fukayama, H., 2011. Functional Incorporation of Sorghum Small Subunit Increases the Catalytic Turnover Rate of Rubisco in Transgenic Rice. Plant Physiology 156, 1603–1611.

Izumi, M., Tsunoda, H., Suzuki, Y., Makino, A., Ishida, H., 2012. RBCS1A and RBCS3B, two major members within the Arabidopsis RBCS multigene family, function to yield sufficient Rubisco content for leaf photosynthetic capacity. J Exp Bot 63, 2159–2170.

Jansing, J., Sack, M., Augustine, S.M., Fischer, R., Bortesi, L., 2019. CRISPR/Cas9-mediated knockout of six glycosyltransferase genes in Nicotiana benthamiana for the production of recombinant proteins lacking β-1,2-xylose and core α-1,3-fucose. Plant Biotechnology Journal 17, 350–361.

Jia, H., Liao, M., Verbelen, J.-P., Vissenberg, K., 2007. Direct creation of marker-free tobacco plants from agroinfiltrated leaf discs. Plant Cell Rep 26, 1961–1965.

Jiang, C.Z., Rodermel, S.R., 1995. Regulation of Photosynthesis during Leaf Development in RbcS Antisense DNA Mutants of Tobacco. Plant Physiology 107, 215–224.

Jiang, C.Z., Rodermel, S.R., Shibles, R.M., 1993. Photosynthesis, Rubisco Activity and Amount, and Their Regulation by Transcription in Senescing Soybean Leaves. Plant Physiology 101, 105–112.

Jiang, F., Doudna, J.A., 2017. CRISPR-Cas9 Structures and Mechanisms. Annu Rev Biophys 46, 505–529.

Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J.A., Charpentier, E., 2012. A Programmable Dual-RNA–Guided DNA Endonuclease in Adaptive Bacterial Immunity. Science 337, 816–821.

John Andrews, T., Whitney, S.M., 2003. Manipulating ribulose bisphosphate carboxylase/oxygenase in the chloroplasts of higher plants. Archives of Biochemistry and Biophysics 414, 159–169.

Jordan, D.B., Ogren, W.L., 1983. Species variation in kinetic properties of ribulose 1,5-bisphosphate carboxylase/oxygenase. Archives of Biochemistry and Biophysics 227, 425–433.

Kanagawa, T., 2003. Bias and artifacts in multitemplate polymerase chain reactions (PCR). Journal of Bioscience and Bioengineering 96, 317–323.

Kanevski, I., Maliga, P., Rhoades, D.F., Gutteridge, S., 1999. Plastome Engineering of Ribulose-1,5-Bisphosphate Carboxylase/Oxygenase in Tobacco to Form a Sunflower Large Subunit and Tobacco Small Subunit Hybrid. Plant Physiology 119, 133–142.

Kapralov, M.V., Filatov, D.A., 2007. Widespread positive selection in the photosynthetic Rubisco enzyme. BMC Evolutionary Biology 7, 73.

Khrebtukova, I., Spreitzer, R.J., 1996. Elimination of the Chlamydomonas gene family that encodes the small subunit of ribulose-1,5-bisphosphate carboxylase/oxygenase. PNAS 93, 13689–13693.

Khumsupan, P., Donovan, S., McCormick, A.J., 2019. CRISPR/Cas in Arabidopsis: overcoming challenges to accelerate improvements in crop photosynthetic efficiencies. Physiologia Plantarum 166, 428–437.

Kim, S., Lee, D.-S., Choi, I.S., Ahn, S.-J., Kim, Y.-H., Bae, H.-J., 2010. Arabidopsis thaliana Rubisco small subunit transit peptide increases the accumulation of Thermotoga maritima endoglucanase Cel5A in chloroplasts of transgenic tobacco plants. Transgenic Res 19, 489–497.

Kloppstech, K., 1985. Diurnal and circadian rhythmicity in the expression of light-induced plant nuclear messenger RNAs. Planta 165, 502–506.

Krapp, A., Chaves, M.M., David, M.M., Rodriques, M.L., Pereira, J.S., Stitt, M., 1994. Decreased ribulose-1,5-bisphosphate carboxylase/oxygenase in transgenic tobacco transformed with 'antisense'rbcS.VIII. Impact on photosynthesis and growth in tobacco growing under extreme high irradiance and high temperature. Plant, Cell & Environment 17, 945–953.

Kubien, D.S., Brown, C.M., Kane, H.J., 2011. Quantifying the amount and activity of Rubisco in leaves. Photosynthesis Research Protocols, 684, 349-362.

Laterre, R., Pottier, M., Remacle, C., Boutry, M., 2017. Photosynthetic Trichomes Contain a Specific Rubisco with a Modified pH-Dependent Activity. Plant Physiology 173, 2110–2120.

Lauerer, M., Saftic, D., Quick, W.P., Labate, C., Fichtner, K., Schulze, E.-D., Rodermel, S.R., Bogorad, L., Stitt, M., 1993. Decreased ribulose-1,5-bisphosphate carboxylase-oxygenase in transgenic tobacco transformed with "antisense" rbcS. Planta 190, 332–345.

Legg, P.D., Chaplin, J.F., Williamson, R.E., 1977. Genetic Diversity in Burley and Flue-cured Tobacco 1. Crop Science 17, 943–947.

Leitch, I.J., Hanson, L., Lim, K.Y., Kovarik, A., Chase, M.W., Clarkson, J.J., Leitch, A.R., 2008. The Ups and Downs of Genome Size Evolution in Polyploid Species of Nicotiana (Solanaceae). Ann Bot 101, 805–814.

Li, J.-F., Aach, J., Norville, J.E., McCormack, M., Zhang, D., Bush, J., Church, G.M., Sheen, J., 2013. Multiplex and homologous recombination-mediated plant genome editing via guide RNA/Cas9. Nat Biotechnol 31, 688–691.

Li, Rui, Li, Ran, Li, X., Fu, D., Zhu, B., Tian, H., Luo, Y., Zhu, H., 2018. Multiplexed CRISPR/Cas9-mediated metabolic engineering of γ-aminobutyric acid levels in Solanum lycopersicum. Plant Biotechnology Journal 16, 415–427.

Lin, M.T., Hanson, M.R., 2018. Red algal Rubisco fails to accumulate in transplastomic tobacco expressing Griffithsia monilis RbcL and RbcS genes. Plant Direct 2, e00045.

Lin, M.T., Occhialini, A., Andralojc, P.J., Parry, M.A.J., Hanson, M.R., 2014. A faster Rubisco with potential to increase photosynthesis in crops. Nature 513, 547–550.

Lin, M.T., Stone, W.D., Chaudhari, V., Hanson, M.R., 2019. Enzyme kinetics of tobacco Rubisco expressed in Escherichia coli varies depending on the small subunit composition. bioRxiv 562223.

Long, S.P., Bernacchi, C.J., 2003. Gas exchange measurements, what can they tell us about the underlying limitations to photosynthesis? Procedures and sources of error. J Exp Bot 54, 2393–2401.

Lowder, L.G., Zhang, D., Baltes, N.J., Paul, J.W., Tang, X., Zheng, X., Voytas, D.F., Hsieh, T.-F., Zhang, Y., Qi, Y., 2015. A CRISPR/Cas9 Toolbox for Multiplexed Plant Genome Editing and Transcriptional Regulation. Plant Physiology 169, 971–985.

Ma, X., Zhang, Q., Zhu, Q., Liu, W., Chen, Yan, Qiu, R., Wang, B., Yang, Z., Li, H., Lin, Y., Xie, Y., Shen, R., Chen, S., Wang, Z., Chen, Yuanling, Guo, J., Chen, L., Zhao, X., Dong, Z., Liu, Y.-G., 2015. A Robust CRISPR/Cas9 System for Convenient, High-Efficiency Multiplex Genome Editing in Monocot and Dicot Plants. Molecular Plant 8, 1274–1284.

Madeira, F., Park, Y.M., Lee, J., Buso, N., Gur, T., Madhusoodanan, N., Basutkar, P., Tivey, A.R.N., Potter, S.C., Finn, R.D., Lopez, R., 2019. The EMBL-EBI search and sequence analysis tools APIs in 2019. Nucleic Acids Res 47, W636–W641.

Makino, A., Shimada, T., Takumi, S., Kaneko, K., Matsuoka, M., Shimamoto, K., Nakano, H., Miyao-Tokutomi, M., Mae, T., Yamamoto, N., 1997. Does Decrease in Ribulose-1,5-Bisphosphate Carboxylase by Antisense RbcS Lead to a Higher N-Use Efficiency of Photosynthesis under Conditions of Saturating CO2 and Light in Rice Plants? Plant Physiology 114, 483–491.

Marshall, B., Biscoe, P.V., 1980. A Model for C3 Leaves Describing the Dependence of Net Photosynthesis on Irradiance. J Exp Bot 31, 29–39.

Masle, J., Hudson, G.S., Badger, M.R., 1993. Effects of Ambient CO2 Concentration on Growth and Nitrogen Use in Tobacco (Nicotiana tabacum) Plants Transformed with an Antisense Gene to the Small Subunit of Ribulose-1,5-Bisphosphate Carboxylase/Oxygenase. Plant Physiology 103, 1075–1088.

Mazur, B.J., Chui, C.-F., 1985. Sequence of a genomic DNA clone for the small subunit of ribulose bis-phosphate carboxylase-oxygenase from tobacco. Nucleic Acids Res 13, 2373–2386.

Miller, A., Schlagnhaufer, C., Spalding, M., Rodermel, S., 2000. Carbohydrate regulation of leaf development: Prolongation of leaf senescence in Rubisco antisense mutants of tobacco. Photosynthesis Research 63, 1–8.

Mitchell, R. a. C., Joyce, P.A., Rong, H., Evans, V.J., Madgwick, P.J., Parry, M. a. J., 2004. Loss of decreased-rubisco phenotype between generations of wheat transformed with antisense and sense rbcS. Annals of Applied Biology 145, 209–216.

Monteith, J., 1991. Plant and crop modelling--A mathematical approach to plant and crop physiology. Agricultural Systems 37, 451–452.

Morineau, C., Bellec, Y., Tellier, F., Gissot, L., Kelemen, Z., Nogué, F., Faure, J.-D., 2017. Selective gene dosage by CRISPR-Cas9 genome editing in hexaploid Camelina sativa. Plant Biotechnology Journal 15, 729–739.

Morita, K., Hatanaka, T., Misoo, S., Fukayama, H., 2016. Identification and expression analysis of non-photosynthetic Rubisco small subunit, OsRbcS1-like genes in plants. Plant Gene 8, 26–31.

Morita, K., Hatanaka, T., Misoo, S., Fukayama, H., 2014. Unusual Small Subunit That Is Not Expressed in Photosynthetic Cells Alters the Catalytic Properties of Rubisco in Rice. Plant Physiology 164, 69–79.

Nagaya, S., Kawamura, K., Shinmyo, A., Kato, K., 2010. The HSP Terminator of Arabidopsis thaliana Increases Gene Expression in Plant Cells. Plant Cell Physiol 51, 328–332.

Nekrasov, V., Staskawicz, B., Weigel, D., Jones, J.D.G., Kamoun, S., 2013. Targeted mutagenesis in the model plant *Nicotiana benthamiana* using Cas9 RNA-guided endonuclease. Nature Biotechnology 31, 691–693.

Norris, S.R., Meyer, S.E., Callis, J., 1993. The intron of Arabidopsis thaliana polyubiquitin genes is conserved in location and is a quantitative determinant of chimeric gene expression. Plant Mol Biol 21, 895–906.

Ogawa, S., Suzuki, Y., Yoshizawa, R., Kanno, K., Makino, A., 2012. Effect of individual suppression of RBCS multigene family on Rubisco contents in rice leaves. Plant, Cell & Environment 35, 546–553.

O'Neal, J.K., Pokalsky, A.R., Kiehne, K.L., Shewmaker, C.K., 1987. Isolation of tobacco SSU genes: characterization of a transcriptionally active pseudogene. Nucleic Acids Res 15, 8661–8677.

Ordon, J., Gantner, J., Kemna, J., Schwalgun, L., Reschke, M., Streubel, J., Boch, J., Stuttmann, J., 2017. Generation of chromosomal deletions in dicotyledonous plants employing a user-friendly genome editing toolkit. The Plant Journal 89, 155–168.

Orr, D.J., Alcântara, A., Kapralov, M.V., Andralojc, P.J., Carmo-Silva, E., Parry, M.A.J., 2016. Surveying Rubisco Diversity and Temperature Response to Improve Crop Photosynthetic Efficiency. Plant Physiology 172, 707–717.

Pan, C., Ye, L., Qin, L., Liu, X., He, Y., Wang, J., Chen, L., Lu, G., 2016. CRISPR/Cas9-mediated efficient and heritable targeted mutagenesis in tomato plants in the first and later generations. Scientific Reports 6, 24765.

Parry, M.A.J., Andralojc, P.J., Scales, J.C., Salvucci, M.E., Carmo-Silva, A.E., Alonso, H., Whitney, S.M., 2013. Rubisco activity and regulation as targets for crop improvement. J Exp Bot 64, 717–730.

Patron, N.J., Orzaez, D., Marillonnet, S., Warzecha, H., Matthewman, C., Youles, M., Raitskin, O., Leveau, A., Farré, G., Rogers, C. *et al.*, 2015. Standards for plant synthetic biology: a common syntax for exchange of DNA parts. New Phytologist 208, 13–19.

Paul, M., Ma, J.K.-C., 2016. Plant-made pharmaceuticals: Leading products and production platforms. IUBMB Life 58–67.

Paulsen, H., Bogorad, L., 1988. Diurnal and Circadian Rhythms in the Accumulation and Synthesis of mRNA for the Light-Harvesting Chlorophyll a/b-Binding Protein in Tobacco. Plant Physiology 88, 1104–1109.

Peterson, B.A., Haak, D.C., Nishimura, M.T., Teixeira, P.J.P.L., James, S.R., Dangl, J.L., Nimchuk, Z.L., 2016. Genome-Wide Assessment of Efficiency and Specificity in CRISPR/Cas9 Mediated Multiple Site Targeting in Arabidopsis. PLOS ONE 11, e0162169.

Pfaffl, M.W., 2001. A new mathematical model for relative quantification in real-time RT–PCR. Nucleic Acids Res 29, e45.

Piechulla, B., Gruissem, W., 1987. Diurnal mRNA fluctuations of nuclear and plastid genes in developing tomato fruits. The EMBO Journal 6, 3593–3599.

Piechulla, B., Pichersky, E., Cashmore, A.R., Gruissem, W., 1986. Expression of nuclear and plastid genes for photosynthesis-specific proteins during tomato fruit development and ripening. Plant Mol Biol 7, 367–376.

Pilgrim, M.L., McClung, C.R., 1993. Differential Involvement of the Circadian Clock in the Expression of Genes Required for Ribulose-1,5-Bisphosphate Carboxylase/Oxygenase Synthesis, Assembly, and Activation in Arabidopsis thaliana. Plant Physiology 103, 553–564.

Pinck, M., Guilley, E., Durr, A., Hoff, M., Pinck, L., Fleck, J., 1984. Complete sequence of one of the mRNAs coding for the small subunit of ribulose bisphosphate carboxylase of Nicotiana sylvestris. Biochimie 66, 539–545.

Porra, R.J., Thompson, W.A., Kriedemann, P.E., 1989. Determination of accurate extinction coefficients and simultaneous equations for assaying chlorophylls a and b extracted with four different solvents: verification of the concentration of chlorophyll standards by atomic absorption spectroscopy. Biochemica et Biophysica Acta (BBA) - Bioenergetics 975, 384–394.

Pottier, M., Gilis, D., Boutry, M., 2018. The Hidden Face of Rubisco. Trends in Plant Science 23, 382–392.

Quick, W.P., Schurr, U., Fichtner, K., Schulze, E.-D., Rodermel, S.R., Bogorad, L., Stitt, M., 1991a. The impact of decreased Rubisco on photosynthesis, growth, allocation and storage in tobacco plants which have been transformed with antisense rbcS. The Plant Journal 1, 51–58.

Quick, W.P., Schurr, U., Fichtner, K., Schulze, E.-D., Rodermel, S.R., Bogorad, L., Stitt, M., 1991b. The impact of decreased Rubisco on photosynthesis, growth, allocation and storage in tobacco plants which have been transformed with antisense rbcS. The Plant Journal 1, 51–58.

Quick, W.P., Schurr, U., Scheibe, R., Schulze, E.D., Rodermel, S.R., Bogorad, L., Stitt, M., 1991c. Decreased ribulose-1,5-bisphosphate carboxylase-oxygenase in transgenic tobacco transformed with “antisense” rbcS : I. Impact on photosynthesis in ambient growth conditions. Planta 183, 542–554.

Raitskin, O., Schudoma, C., West, A., Patron, N.J., 2019. Comparison of efficiency and specificity of CRISPR-associated (Cas) nucleases in plants: An expanded toolkit for precision genome engineering. PLOS ONE 14, e0211598.

Rodermel, S., Haley, J., Jiang, C.Z., Tsai, C.H., Bogorad, L., 1996. A mechanism for intergenomic integration: abundance of ribulose bisphosphate carboxylase small-subunit protein influences the translation of the large-subunit mRNA. PNAS 93, 3881–3885.

Rodermel, S.R., Abbott, M.S., Bogorad, L., 1988. Nuclear-organelle interactions: Nuclear antisense gene inhibits ribulose bisphosphate carboxylase enzyme levels in transformed tobacco plants. Cell 55, 673–681.

Ruf, S., Forner, J., Hasse, C., Kroop, X., Seeger, S., Schollbach, L., Schadach, A., Bock, R., 2019. High-efficiency generation of fertile transplastomic Arabidopsis plants. Nat. Plants 5, 282–289.

Sappl, P.G., Heazlewood, J.L., Millar, A.H., 2004. Untangling multi-gene families in plants by integrating proteomics into functional genomics. Phytochemistry, Proteomics 1 65, 1517–1530.

Sawbridge, T.I., López-Juez, E., Knight, M.R., Jenkins, G.I., 1993. A blue-light photoreceptor mediates the fluence-rate-dependent expression of genes encoding the small subunit of ribulose 1,5-bisphosphate carboxylase/oxygenase in light-grown Phaseolus vulgaris primary leaves. Planta 192, 1–8.

Sawchuk, M.G., Donner, T.J., Head, P., Scarpella, E., 2008. Unique and Overlapping Expression Patterns among Members of Photosynthesis-Associated Nuclear Gene Families in Arabidopsis. Plant Physiology 148, 1908–1924.

Schäfer, E., Haupt, W., 1983. Blue-Light Effects in Phytochrome-Mediated Responses, in: Shropshire, W., Mohr, H. (Eds.), Photomorphogenesis, Encyclopedia of Plant Physiology. Springer Berlin Heidelberg, pp. 723–744.

Schatz, M.C., Witkowski, J., McCombie, W.R., 2012. Current challenges in de novo plant genome sequencing and assembly. Genome Biology 13, 243.

Schmidt, G.W., Delaney, S.K., 2010. Stable internal reference genes for normalization of real-time RT-PCR in tobacco (Nicotiana tabacum) during development and abiotic stress. Mol Genet Genomics 283, 233–241.

Schneider, C.A., Rasband, W.S., Eliceiri, K.W., 2012. NIH Image to ImageJ: 25 years of image analysis. Nat Methods 9, 671–675.

Sharkey, T.D., 1985. Photosynthesis in intact leaves of C3 plants: Physics, physiology and rate limitations. Bot. Rev 51, 53–105

Sharwood, R.E., Caemmerer, S. von, Maliga, P., Whitney, S.M., 2008. The Catalytic Properties of Hybrid Rubisco Comprising Tobacco Small and Sunflower Large Subunits Mirror the Kinetically Equivalent Source Rubiscos and Can Support Tobacco Growth. Plant Physiology 146, 83–96.

Sheen, J.-Y., Bogorad, L., 1986. Expression of the ribulose-1,5-bisphosphate carboxylase large subunit gene and three small subunit genes in two cell types of maize leaves. EMBO J 5, 3417–3422.

Shen, L., Hua, Y., Fu, Y., Li, J., Liu, Q., Jiao, X., Xin, G., Wang, J., Wang, X., Yan, C., Wang, K., 2017. Rapid generation of genetic diversity by multiplex CRISPR/Cas9 genome editing in rice. Sci. China Life Sci. 60, 506–515.

Shmakov, S., Smargon, A., Scott, D., Cox, D., Pyzocha, N., Yan, W., Abudayyeh, O.O., Gootenberg, J.S., Makarova, K.S., Wolf, Y.I., Severinov, K., Zhang, F., Koonin, E.V., 2017. Diversity and evolution of class 2 CRISPR–Cas systems. Nature Reviews Microbiology 15, 169–182.

Sierro, N., Battey, J.N.D., Ouadi, S., Bakaher, N., Bovet, L., Willig, A., Goepfert, S., Peitsch, M.C., Ivanov, N.V., 2014. The tobacco genome sequence and its comparison with those of tomato and potato. Nature Communications 5, 3833.

Silverthorne, J., Wimpee, C.F., Yamada, T., Rolfe, S.A., Tobin, E.M., 1990. Differential expression of individual genes encoding the small subunit of ribulose-1,5-bisphosphate carboxylase in Lemna gibba. Plant Mol Biol 15, 49–58. 3

Simpson, E., Cooke, R.J., Davies, D.D., 1981. Measurement of protein degradation in leaves of Zea mays using acetic anhydride and tritiated water. Plant Physiology 67, 1214-1219.

Smith, C.J.S., Watson, C.F., Bird, C.R., Ray, J., Schuch, W., Grierson, D., 1990. Expression of a truncated tomato polygalacturonase gene inhibits expression of the endogenous gene in transgenic plants. Molec. Gen. Genet. 224, 477–481.

Smith, H., 2000. Phytochromes and light signal perception by plants—an emerging synthesis. Nature 407, 585–591.

Sparkes, I.A., Runions, J., Kearns, A., Hawes, C., 2006. Rapid, transient expression of fluorescent fusion proteins in tobacco plants and generation of stably transformed plants. Nature Protocols 1

Spreitzer, R.J., 2003. Role of the small subunit in ribulose-1,5-bisphosphate carboxylase/oxygenase. Archives of Biochemistry and Biophysics 414, 141–149.

Spreitzer, R.J., Salvucci, M.E., 2002. RUBISCO: Structure, Regulatory Interactions, and Possibilities for a Better Enzyme. Annual Review of Plant Biology 53, 449–475.

Srivastava, V., Underwood, J.L., Zhao, S., 2017. Dual-targeting by CRISPR/Cas9 for precise excision of transgenes from rice genome. Plant Cell Tiss Organ Cult 129, 153–160.

Stayton, M.M., Brosio, P., Dunsmuir, P., 1989. Photosynthetic Genes of Petunia (Mitchell) Are Differentially Expressed during the Diurnal Cycle. Plant Physiology 89, 776–

Stitt, M., Quick, W.P., Schurr, U., Schulze, E.-D., Rodermel, S.R., Bogorad, L., 1991. Decreased ribulose-1,5-bisphosphate carboxylase-oxygenase in transgenic tobacco transformed with ‘antisense’ rbcS. Planta 183, 555–566.

Stitt, M., Schulze, D., 1994. Does Rubisco control the rate of photosynthesis and plant growth? An exercise in molecular ecophysiology. Plant Cell and Environment 17, 465–487.

Sugita, M., Gruissem, W., 1987. Developmental, organ-specific, and light-dependent expression of the tomato ribulose-1,5-bisphosphate carboxylase small subunit gene family. PNAS 84, 7104–7108.

Sugita, M., Manzara, T., Pichersky, E., Cashmore, A., Gruissem, W., 1987. Genomic organization, sequence analysis and expression of all five genes encoding the small subunit of ribulose-1,5-bisphosphate carboxylase/oxygenase from tomato. Molec Gen Genet 209, 247–256.

Sun, B., Zheng, A., Jiang, M., Xue, S., Yuan, Q., Jiang, L., Chen, Q., Li, M., Wang, Y., Zhang, Y., Luo, Y., Wang, X., Zhang, F., Tang, H., 2018. CRISPR/Cas9-mediated mutagenesis of homologous genes in Chinese kale. Sci Rep 8, 1–10.

Suzuki, Y., Nakabayashi, K., Yoshizawa, R., Mae, T., Makino, A., 2009. Differences in Expression of the RBCS Multigene Family and Rubisco Protein Content in Various Rice Plant Tissues at Different Growth Stages. Plant Cell Physiol 50, 1851–1855.

Svab, Z., Maliga, P., 1993. High-frequency plastid transformation in tobacco by selection for a chimeric aadA gene. Proc Natl Acad Sci U S A 90, 913–917.

Tabita, F.R., Satagopan, S., Hanson, T.E., Kreel, N.E., Scott, S.S., 2008. Distinct form I, II, III, and IV Rubisco proteins from the three kingdoms of life provide clues about Rubisco evolution and structure/function relationships. J Exp Bot 59, 1515–1524.

Tang, X., Ren, Q., Yang, L., Bao, Y., Zhong, Z., He, Y., Liu, S., Qi, C., Liu, B., Wang, Y., Sretenovic, S., Zhang, Yingxiao, Zheng, X., Zhang, T., Qi, Y., Zhang, Yong, 2019. Single transcript unit CRISPR 2.0 systems for robust Cas9 and Cas12a mediated plant genome editing. Plant Biotechnology Journal 17, 1431–1445.

Taylor, T.C., Andersson, I., 1997. The structure of the complex between rubisco and its natural substrate ribulose 1,5-bisphosphate11Edited by R. Huber. Journal of Molecular Biology 265, 432–444.

Tieman, D.M., Taylor, M.G., Ciardi, J.A., Klee, H.J., 2000. The tomato ethylene receptors NR and LeETR4 are negative regulators of ethylene response and exhibit functional compensation within a multigene family. PNAS 97, 5663–5668.

Tsai, C.H., Miller, A., Spalding, M., Rodermel, S., 1997. Source Strength Regulates an Early Phase Transition of Tobacco Shoot Morphogenesis. Plant Physiology 115, 907–914.

Tsai, Y.-C.C., Mueller-Cajar, O., Saschenbrecker, S., Hartl, F.U., Hayer-Hartl, M., 2012. Chaperonin Cofactors, Cpn10 and Cpn20, of Green Algae and Plants Function as Hetero-oligomeric Ring Complexes. J. Biol. Chem. 287, 20471–20481.

von Caemmerer, S., Lawson, T., Oxborough, K., Baker, N.R., Andrews, T.J., Raines, C.A., 2004. Stomatal conductance does not correlate with photosynthetic capacity in transgenic tobacco with reduced amounts of Rubisco. J Exp Bot 55, 1157–1166.

Voytas, D.F., 2013. Plant Genome Engineering with Sequence-Specific Nucleases. Annual Review of Plant Biology 64, 327–350.

Wang, W., Pan, Q., He, F., Akhunova, A., Chao, S., Trick, H., Akhunov, E., 2018. Transgenerational CRISPR-Cas9 Activity Facilitates Multiplex Gene Editing in Allopolyploid Wheat. The CRISPR Journal 1, 65–74.

Wang, Y., Cheng, X., Shan, Q., Zhang, Y., Liu, J., Gao, C., Qiu, J.-L., 2014. Simultaneous editing of three homoeoalleles in hexaploid bread wheat confers heritable resistance to powdery mildew. Nature Biotechnology 32, 947–951.

Wanner, L.A., Gruissem, W., 1991. Expression dynamics of the tomato rbcS gene family during development. The Plant Cell 3, 1289–1303.

Weeks, K.E., Chuzhanova, N.A., Donnison, I.S., Scott, I.M., 2007. Evolutionary hierarchies of conserved blocks in 5'-noncoding sequences of dicot rbcS genes. BMC Evolutionary Biology 7, 51.

Wehmeyer, B., Cashmore, A.R., Schäfer, E., 1990. Photocontrol of the Expression of Genes Encoding Chlorophyll a/b Binding Proteins and Small Subunit of Ribulose-1,5-Bisphosphate Carboxylase in Etiolated Seedlings of Lycopersicon esculentum (L.) and Nicotiana tabacum (L.). Plant Physiology 93, 990–997.

Wendel, J.F., 2000. Genome evolution in polyploids, in: Doyle, J.J., Gaut, B.S. (Eds.), Plant Molecular Evolution. Springer Netherlands, Dordrecht, pp. 225–249.

Whitney, S.M., Andrews, T.J., 2003. Photosynthesis and Growth of Tobacco with a Substituted Bacterial Rubisco Mirror the Properties of the Introduced Enzyme. Plant Physiology 133, 287–294.

Whitney, S.M., Birch, R., Kelso, C., Beck, J.L., Kapralov, M.V., 2015. Improving recombinant Rubisco biogenesis, plant photosynthesis and growth by coexpressing its ancillary RAF1 chaperone. PNAS 112, 3564–3569.

Whitney, S.M., Caemmerer, S. von, Hudson, G.S., Andrews, T.J., 1999. Directed Mutation of the Rubisco Large Subunit of Tobacco Influences Photorespiration and Growth. Plant Physiology 121, 579–588.

Whitney, S.M., Houtz, R.L., Alonso, H., 2011a. Advancing Our Understanding and Capacity to Engineer Nature's CO2-Sequestering Enzyme, Rubisco. Plant Physiology 155, 27–35.

Whitney, S.M., Sharwood, R.E., 2008. Construction of a tobacco master line to improve Rubisco engineering in chloroplasts. J Exp Bot 59, 1909–1921.

Whitney, S.M., Sharwood, R.E., Orr, D., White, S.J., Alonso, H., Galmés, J., 2011b. Isoleucine 309 acts as a C4 catalytic switch that increases ribulose-1,5-bisphosphate carboxylase/oxygenase (rubisco) carboxylation rate in Flaveria. Proc Natl Acad Sci U S A 108, 14688–14693.

Wilson, R.H., Alonso, H., Whitney, S.M., 2016. Evolving *Methanococcoides burtonii* archaeal Rubisco for improved photosynthesis and plant growth. Scientific Reports 6, 22284.

Wolter, F.P., Fritz, C.C., Willmitzer, L., Schell, J., Schreier, P.H., 1988. rbcS genes in Solanum tuberosum: conservation of transit peptide and exon shuffling during evolution. PNAS 85, 846–850.

Wostrikoff, K., Stern, D., 2007. Rubisco large-subunit translation is autoregulated in response to its assembly state in tobacco chloroplasts. PNAS 104, 6466–6471.

Xie, K., Yang, Y., 2013. RNA-Guided Genome Editing in Plants Using a CRISPR–Cas System. Molecular Plant 6, 1975–1983.

Xie, X., Qin, G., Si, P., Luo, Z., Gao, J., Chen, X., Zhang, J., Wei, P., Xia, Q., Lin, F., Yang, J., 2017. Analysis of Nicotiana tabacum PIN genes identifies NtPIN4 as a key regulator of axillary bud growth. Physiologia Plantarum 160, 222–239.

Xing, H.-L., Dong, L., Wang, Z.-P., Zhang, H.-Y., Han, C.-Y., Liu, B., Wang, X.-C., Chen, Q.-J., 2014. A CRISPR/Cas9 toolkit for multiplex genome editing in plants. BMC Plant Biology 14, 327.

Xu, R.-F., Li, H., Qin, R.-Y., Li, J., Qiu, C.-H., Yang, Y.-C., Ma, H., Li, L., Wei, P.-C., Yang, J.-B., 2015. Generation of inheritable and "transgene clean" targeted genome-modified rice in later generations using the CRISPR/Cas9 system. Scientific Reports

Yoon, M., Putterill, J.J., Ross, G.S., Laing, W.A., 2001. Determination of the Relative Expression Levels of Rubisco Small Subunit Genes in Arabidopsis by Rapid Amplification of cDNA Ends. Analytical Biochemistry 291, 237–244.

Yu, Q., Lutz, K.A., Maliga, P., 2017. Efficient Plastid Transformation in Arabidopsis. Plant Physiology 175, 186–193.

Zhang, C., Liu, C., Weng, J., Cheng, B., Liu, F., Li, X., Xie, C., 2017. Creation of targeted inversion mutations in plants using an RNA-guided endonuclease. The Crop Journal 5, 83–88.

Zhang, Hui, Zhang, J., Wei, P., Zhang, B., Gou, F., Feng, Z., Mao, Y., Yang, L., Zhang, Heng, Xu, N., Zhu, J.-K., 2014. The CRISPR/Cas9 system produces specific and homozygous targeted gene editing in rice in one generation. Plant Biotechnology Journal 12, 797–807.

Zhang, X.-H., Webb, J., Huang, Y.-H., Lin, L., Tang, R.-S., Liu, A., 2011. Hybrid Rubisco of tomato large subunits and tobacco small subunits is functional in tobacco plants. Plant Science 180, 480–488.

Zhang, Y., Liang, Z., Zong, Y., Wang, Y., Liu, J., Chen, K., Qiu, J.-L., Gao, C., 2016. Efficient and transgene-free genome editing in wheat through transient expression of CRISPR/Cas9 DNA or RNA. Nature Communications 7, 12617.

Zhang, Z., Mao, Y., Ha, S., Liu, W., Botella, J.R., Zhu, J.-K., 2016. A multiplex CRISPR/Cas9 platform for fast and efficient editing of multiple genes in Arabidopsis. Plant Cell Rep 35, 1519–1533.

Zhao, Y., Zhang, C., Liu, W., Gao, W., Liu, C., Song, G., Li, W.-X., Mao, L., Chen, B., Xu, Y., Li, X., Xie, C., 2016. An alternative strategy for targeted gene replacement in plants using a dual-sgRNA/Cas9 design. Scientific Reports 6, 23890.

Zhou, H., Liu, B., Weeks, D.P., Spalding, M.H., Yang, B., 2014. Large chromosomal deletions and heritable small genetic changes induced by CRISPR/Cas9 in rice. Nucleic Acids Res 42, 10903–10914.

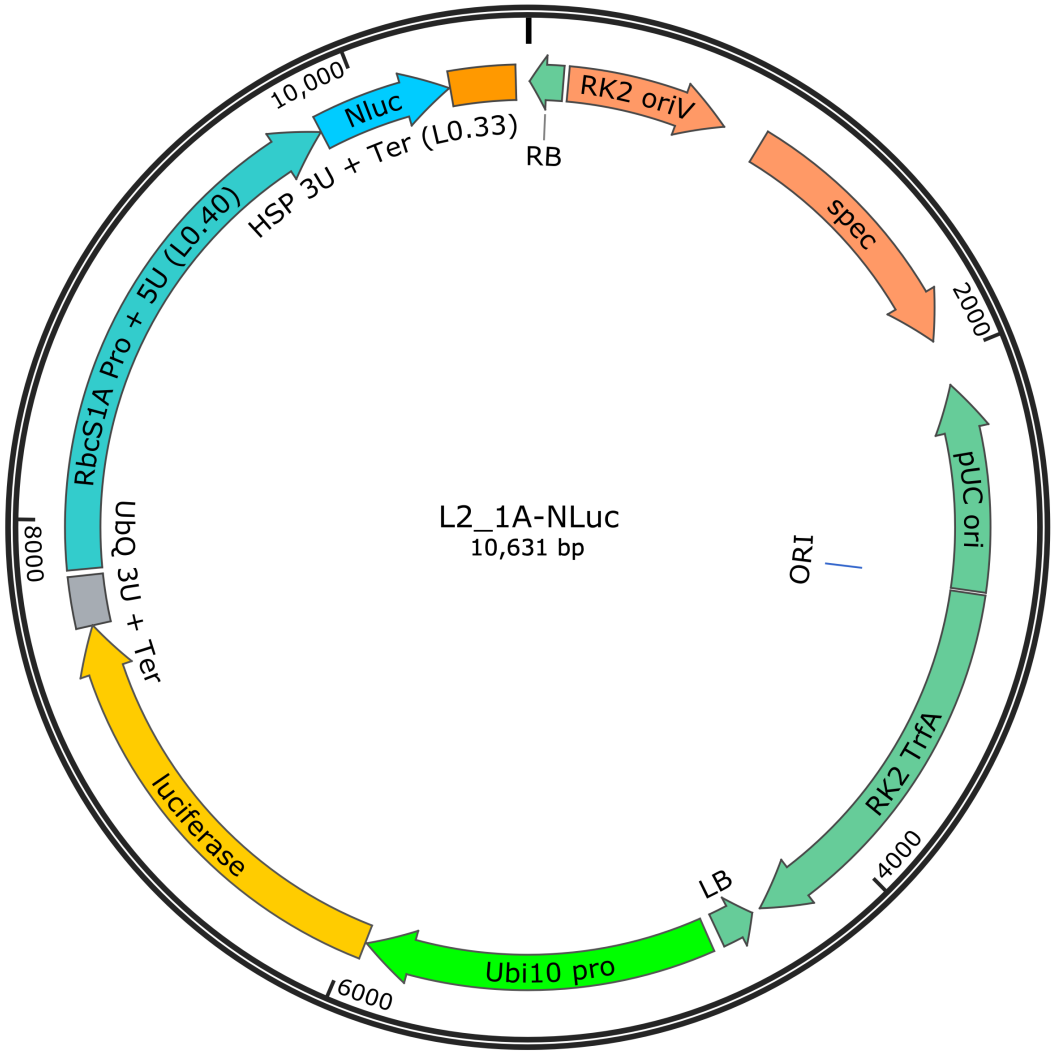
Zhu, X.-G., Long, S.P., Ort, D.R., 2010. Improving Photosynthetic Efficiency for Greater Yield. Annual Review of Plant Biology 61, 235–261.

Zhu, X.-G., Long, S.P., Ort, D.R., 2008. What is the maximum efficiency with which photosynthesis can convert solar energy into biomass? Current Opinion in Biotechnology, Food biotechnology / Plant biotechnology 19, 153–159.

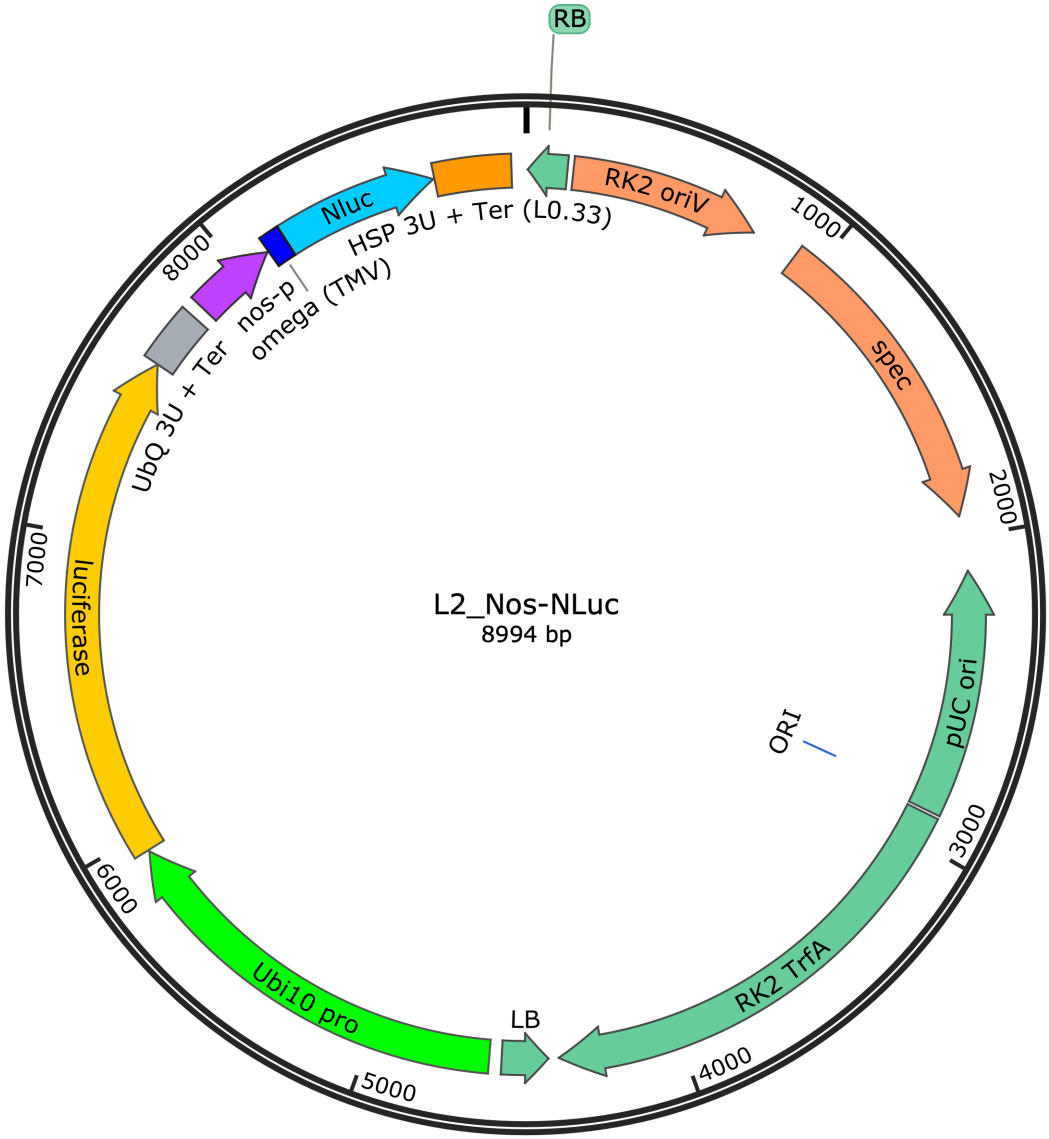
Appendices

Appendix A: Plasmid maps

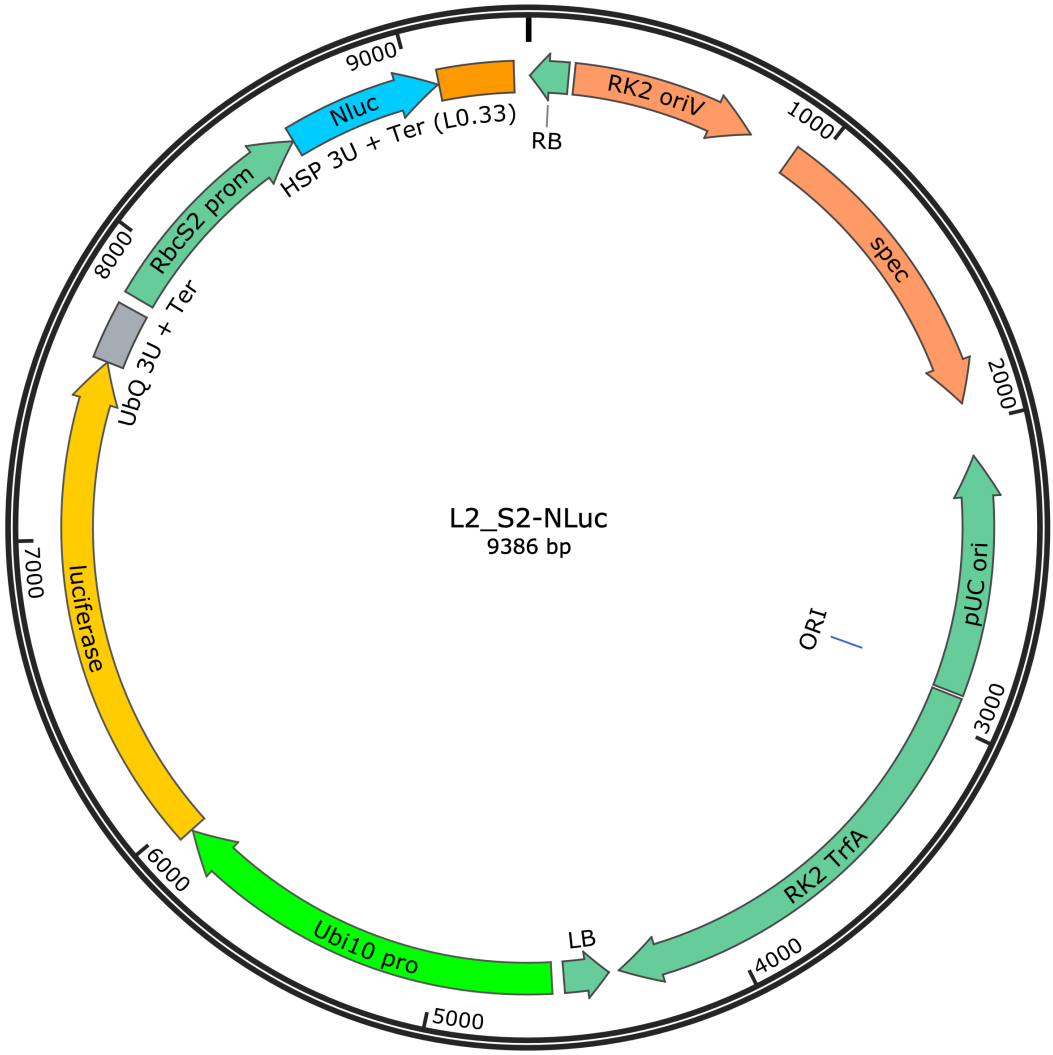
Vectors for transient assays



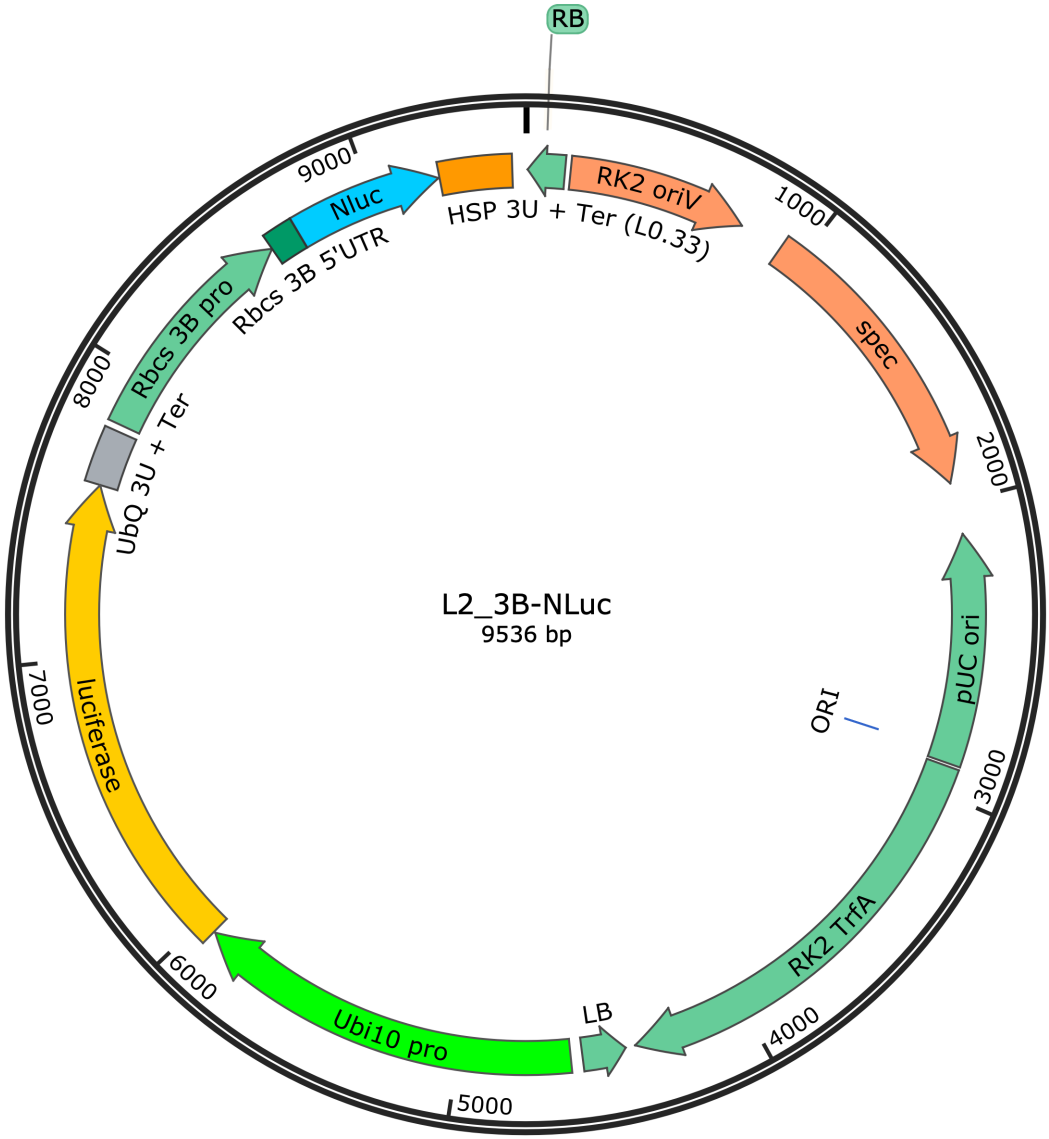
Appendix A Figure 1. L2_1A_NLuc vector (Table 2–4) used for dual luciferase assay in protoplasts. The following parts were assembled into a Level M acceptor vector (pAGM8031): Ubi10 pro, promoter and 5' UTR, *ubiquitin 10* (*A. thaliana*) (pICSL12015); luciferase, luciferase coding sequence (*Photinus pyralis*) (pICSL50006); UBQ 3U + Ter, terminator and 3' UTR, *ubiquitin 5* (*A. thaliana*); RbcS1A Pro + 5U, promoter and 5' UTR, *rbcS1A* (*A. thaliana*); NLuc, nano luciferase coding sequence; HSP 3U + Ter, terminator and 3' UTR, *heat shock protein 18.2* (*A. thaliana*); RB, right border; LB, left border.



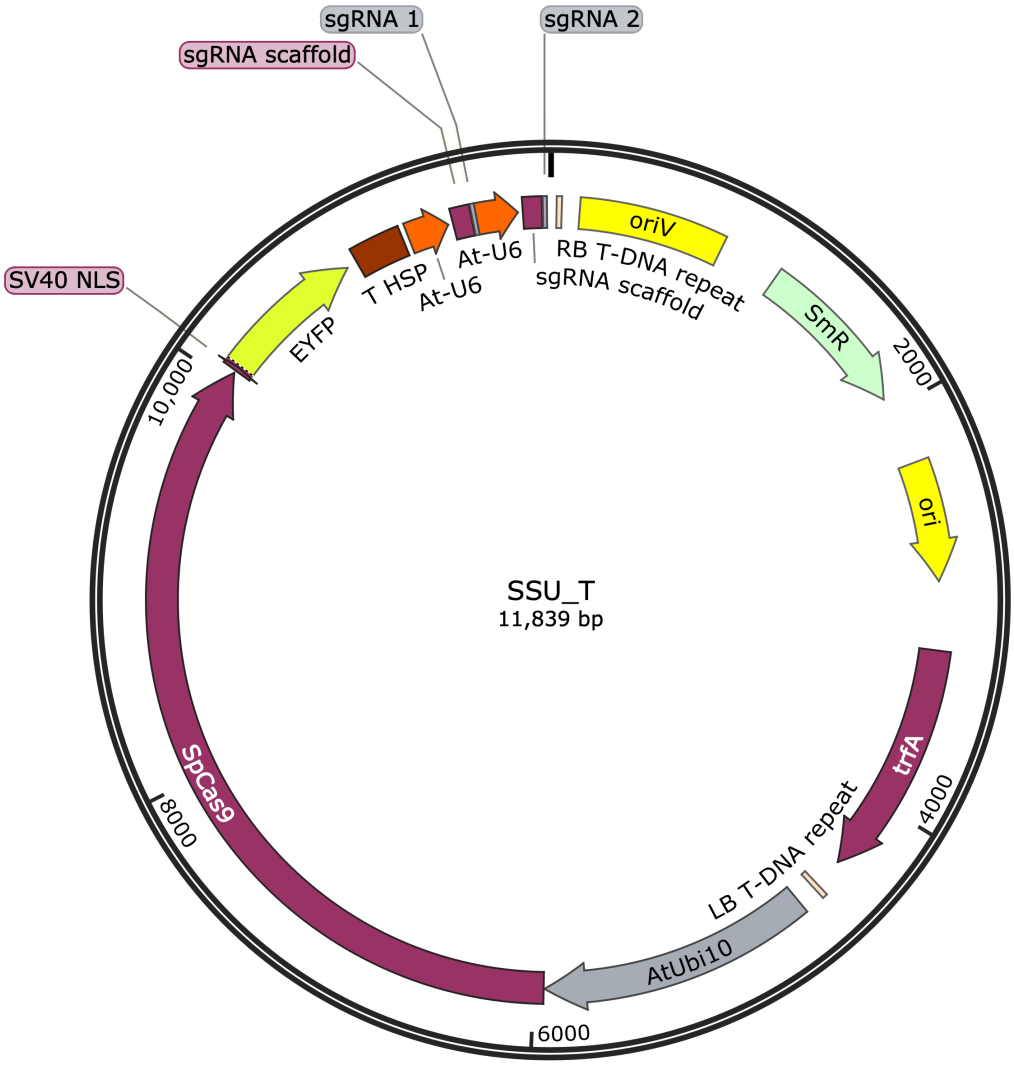
Appendix A Figure 2. L2_Nos_NLuc vector (Table 2–4) used for dual luciferase assay in protoplasts. The following parts were assembled into a Level M acceptor vector (pAGM8031): Ubi10 pro, promoter and 5' UTR, *ubiquitin 10* (*A.thaliana*) (pICSL12015); luciferase, luciferase coding sequence (*Photinus pyralis*) (pICSL50006); UBQ 3U + Ter, terminator and 3' UTR, *ubiquitin 5* (*A.thaliana*); nos-p, promoter, *nopaline synthase* (*A.tumefacians*) (pICH87633); omega (TMV), 5' UTR (tobacco mosaic virus); NLuc, nano luciferase coding sequence; HSP 3U + Ter, terminator and 3' UTR, *heat shock protein 18.2* (*A.thaliana*); RB, right border; LB, left border.



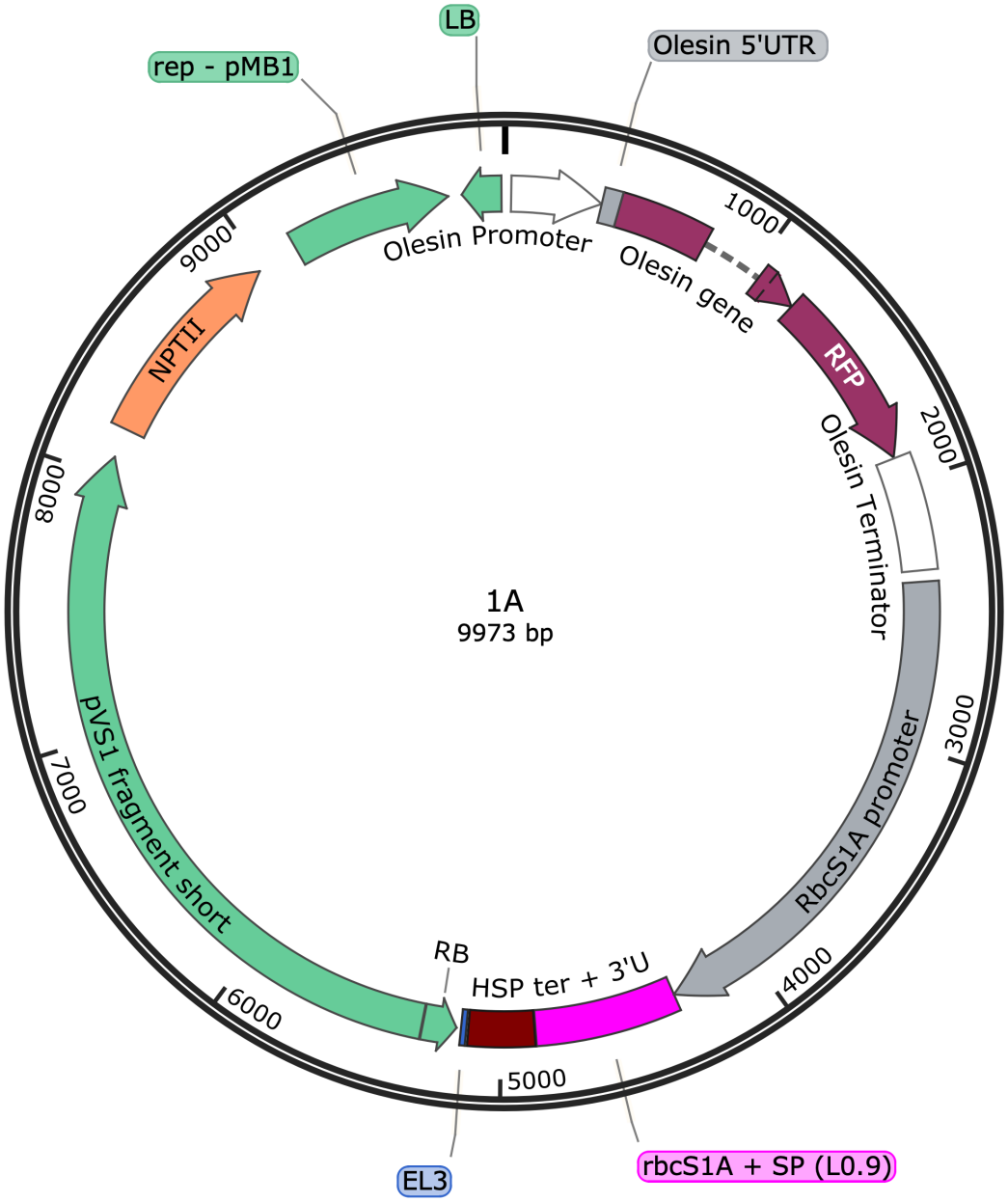
Appendix A Figure 3. L2_S2_NLuc vector (Table 2–4) used for dual luciferase assay in protoplasts. The following parts were assembled into a Level M acceptor vector (pAGM8031): Ubi10 pro, promoter and 5' UTR, *ubiquitin 10* (*A. thaliana*) (pICSL12015); luciferase, luciferase coding sequence (*Photinus pyralis*) (pICSL50006); UBQ 3U + Ter, terminator and 3' UTR, *ubiquitin 5* (*A. thaliana*); RbcS2 prom, promoter and 5' UTR, *rbcS2* (*S. lycopersicum*) (pICH71301); NLuc, nano luciferase coding sequence; HSP 3U + Ter, terminator and 3' UTR, *heat shock protein 18.2* (*A. thaliana*); RB, right border; LB, left border.



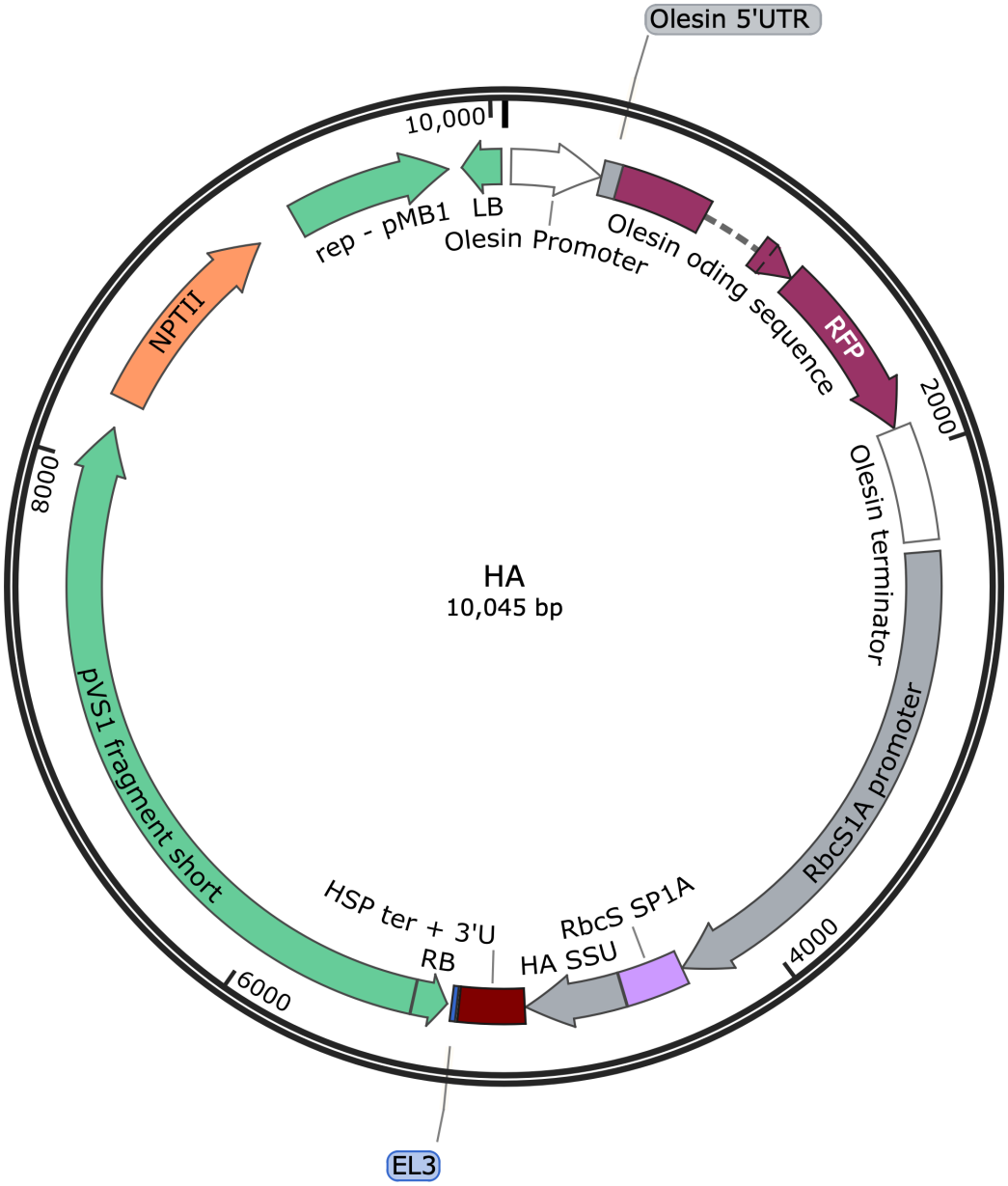
Appendix A Figure 4. L2_3B_NLuc vector (Table 2–4) used for dual luciferase assay in protoplasts. The following parts were assembled into a Level M acceptor vector (pAGM8031): Ubi10 pro, promoter and 5' UTR, *ubiquitin 10* (*A. thaliana*) (pICSL12015); luciferase, luciferase coding sequence (*Photinus pyralis*) (pICSL50006); UBQ 3U + Ter, terminator and 3' UTR, *ubiquitin 5* (*A. thaliana*); RbcS3B prom, promoter and 5' UTR, *rbcS3B* (*A. thaliana*) (pICH45244); NLuc, nano luciferase coding sequence; HSP 3U + Ter, terminator and 3' UTR, *heat shock protein 18.2* (*A. thaliana*); RB, right border; LB, left border.



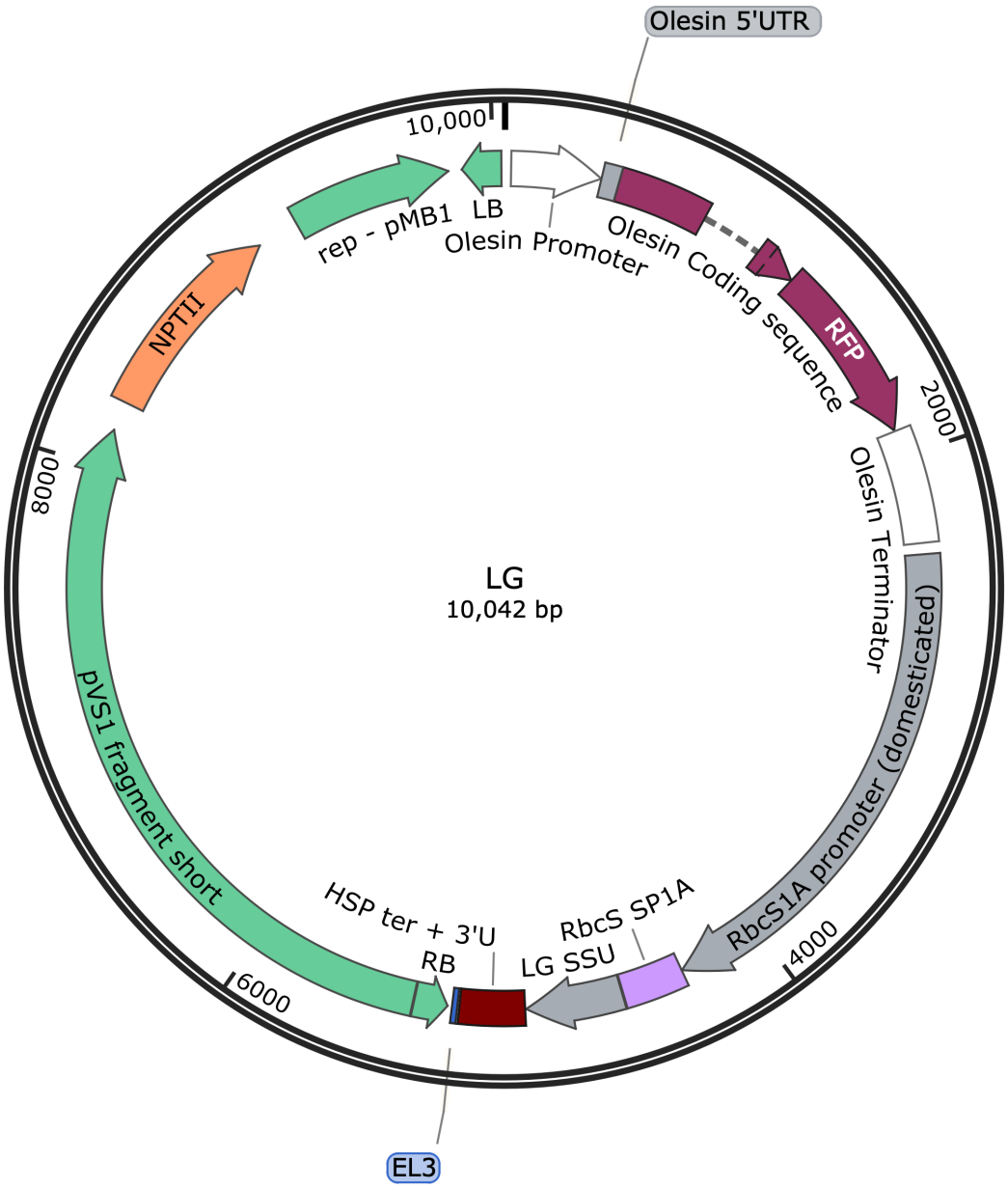
Appendix A Figure 5. SSU_T vector (Table 2–4) used for CRISPR/Cas9 transient assays. The following parts were assembled into a Level M acceptor vector (pAGM8031): Ubi10 pro, promoter and 5' UTR, *ubiquitin 10* (*A. thaliana*) (pICSL12015); spCas9, *cas9* coding sequence (*S. pyogenes*) (pICSL90005); SV40 NLS, nuclear localisation signal (simian virus 40); eYFP, n-terminal tag (pICSL50017); T HSP, terminator and 3' UTR, *heat shock protein 18.2* (*A. thaliana*); At-U6, promoter, *U6-26* (*A. thaliana*); sgRNA scaffold, sgRNA scaffold sequence (pICSL90010).



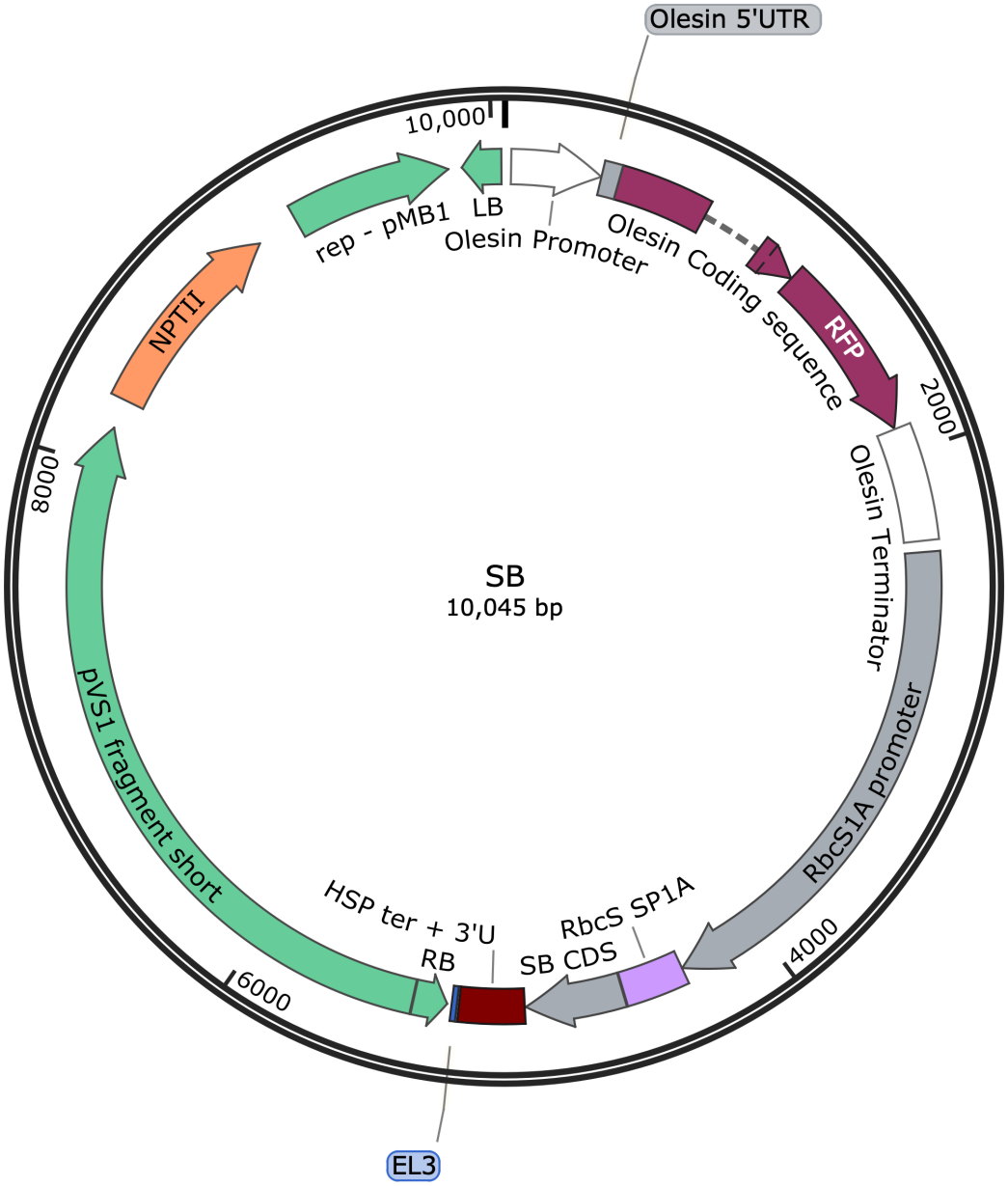
Appendix A Figure 6. 1A vector (Table 2–5) used to transform the Arabidopsis *1a3b* mutant. The following parts were assembled into a Level 2 acceptor vector (pAGM4723): FAST selection marker (pICSL11015) containing promoter, 5' UTR, coding sequence, and terminator from *olesin* (*A. thaliana*) and RFP c-terminal tag; RbcS1A promoter, promoter and 5' UTR, *rbcS1A* (*A. thaliana*); rbcS1A + SP, coding sequence and signal peptide, *rbcS1A* (*A. thaliana*); HSP 3U + Ter, terminator and 3' UTR, *heat shock protein 18.2* (*A. thaliana*); RB, right border; LB, left border.



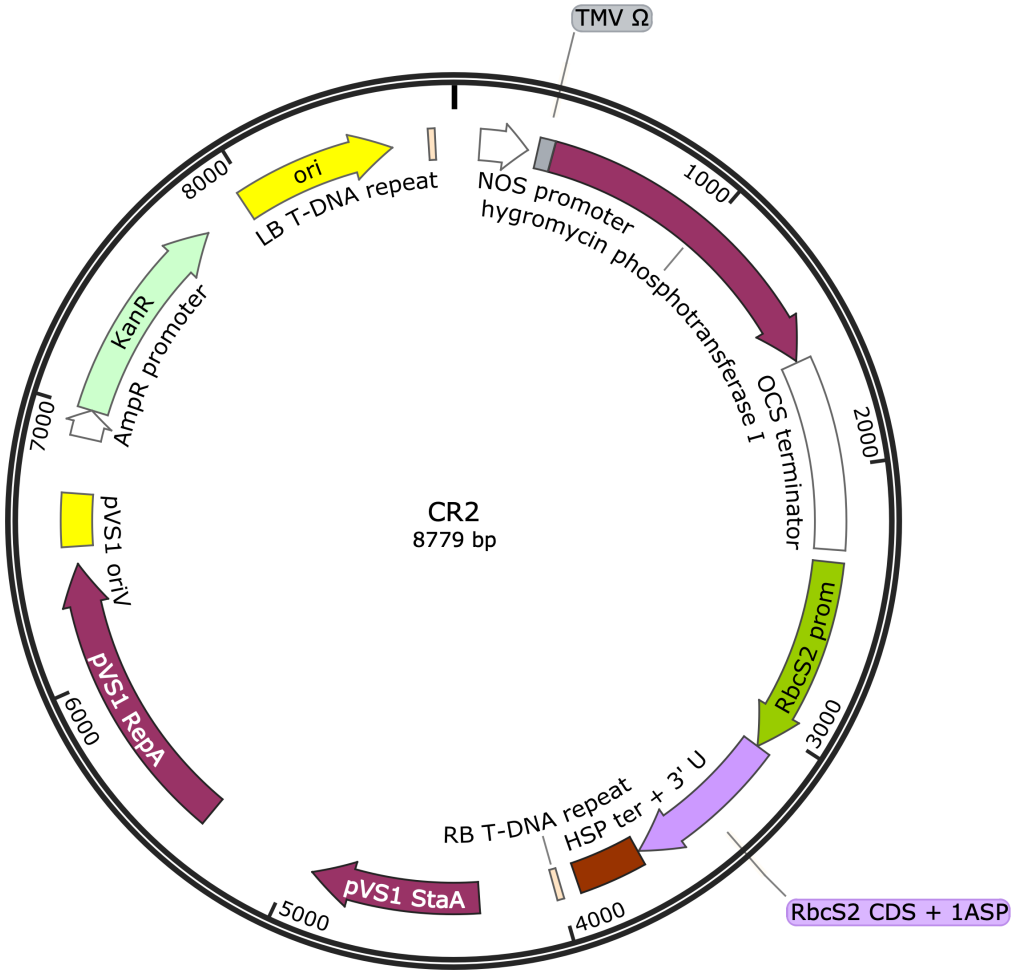
Appendix A Figure 7. HA vector (Table 2–5) used to transform the Arabidopsis *1a3b* mutant. The following parts were assembled into a Level 2 acceptor vector (pAGM4723): FAST selection marker (pICSL11015) containing promoter, 5' UTR, coding sequence, and terminator from *olesin* (*A. thaliana*) and RFP c-terminal tag; RbcS1A promoter, promoter and 5' UTR, *rbcS1A* (*A. thaliana*); rbcS1A SP1A, signal peptide, *rbcS1A* (*A. thaliana*); HA SSU, small subunit coding sequence (*H. annus*) (Table 2-7); HSP 3U + Ter, terminator and 3' UTR, *heat shock protein 18.2* (*A. thaliana*); RB, right border; LB, left border.



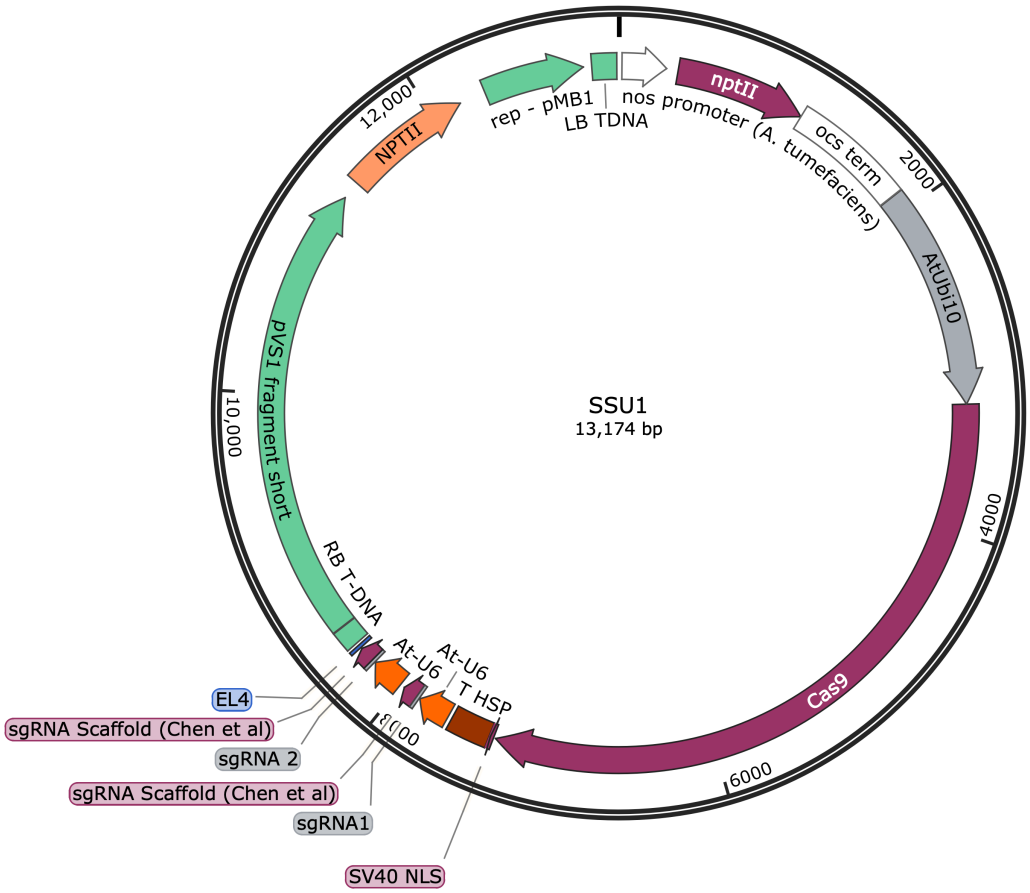
Appendix A Figure 8. LG vector (Table 2–5) used to transform the Arabidopsis *1a3b* mutant. The following parts were assembled into a Level 2 acceptor vector (pAGM4723): FAST selection marker (pICSL11015) containing promoter, 5' UTR, coding sequence, and terminator from *olesin* (*A. thaliana*) and RFP c-terminal tag; RbcS1A promoter, promoter and 5' UTR, *rbcS1A* (*A. thaliana*); rbcS1A SP1A, signal peptide, *rbcS1A* (*A. thaliana*); HA SSU, small subunit coding sequence (*L. gibertii*) (Table 2-7); HSP 3U + Ter, terminator and 3' UTR, *heat shock protein 18.2* (*A. thaliana*); RB, right border; LB, left border.



Appendix A Figure 9. SB vector (Table 2–5) used to transform the Arabidopsis *1a3b* mutant. The following parts were assembled into a Level 2 acceptor vector (pAGM4723): FAST selection marker (pICSL11015) containing promoter, 5' UTR, coding sequence, and terminator from *olesin* (*A. thaliana*) and RFP c-terminal tag; RbcS1A promoter, promoter and 5' UTR, *rbcS1A* (*A. thaliana*); rbcS1A SP1A, signal peptide, *rbcS1A* (*A. thaliana*); HA SSU, small subunit coding sequence (*S. bicolor*) (Table 2-7); HSP 3U + Ter, terminator and 3' UTR, *heat shock protein 18.2* (*A. thaliana*); RB, right border; LB, left border.



Appendix A Figure 10. CR vector (Table 2–5) used to transform wild-type tobacco. The following parts were assembled into a Level 2 acceptor vector (pAGM4723): hygromycin phosphotransferase I (pICSL80036) with nos promoter, *nopaline synthase* (*A. tumefacians*) (pICH87633) and ocs terminator, *octopine synthase* (*A. tumefaciens*) (pICH41432);); RbcS2 prom, promoter and 5' UTR, *rbcS2* (*S. lycopersicum*) (pICH71301); *rbcS1A* signal peptide (*A. thaliana*) and *rbcS2* coding sequence (*C. reinhardtti*); HSP 3U + Ter, terminator and 3' UTR, *heat shock protein 18.2* (*A. thaliana*); RB, right border; LB, left border.

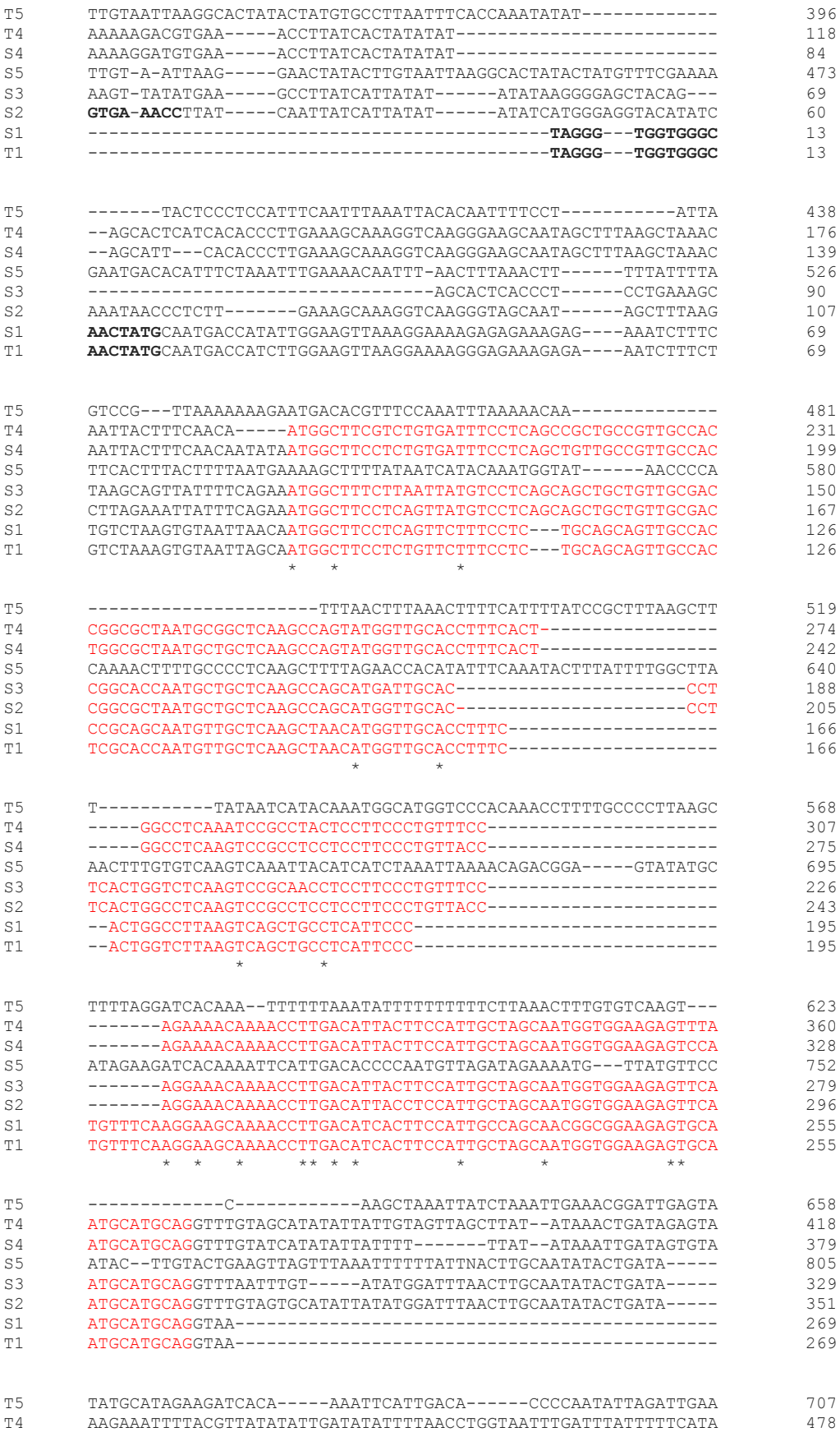


Appendix A Figure 11. SSU1 vector (Table 2–5) used for CRISPR/Cas9 editing in tobacco. The following parts were assembled into a Level 2 acceptor vector (pAGM4723): nos promoter, *nopaline synthase* promoter (*A. tumefaciens*); nptII, kanamycin resistance (*E.coli*); ocs term, *octopine synthase* terminator (*A. tumefaciens*); Ubi10 pro, promoter and 5' UTR, *ubiquitin 10* (*A. thaliana*) (pICSL12015); spCas9, *cas9* coding sequence (*S. pyogenes*) (pICSL90004); T HSP, terminator and 3' UTR, *heat shock protein 18.2* (*A. thaliana*); At-U6, promoter, *U6-26* (*A. thaliana*); sgRNA scaffold, sgRNA scaffold sequence (pICSL90010).

Appendix B: Multiple sequence alignments (MSAs)

Full-length genomic DNA sequences

```
T5      ------------------------------------------------CCGTATGTGCA        11
T4      ------------------------------------------------------------      0
S4      ------------------------------------------------------------      0
S5      AACGGCTACCATTCCTCTCATCTTAAGATGAGGTTTCCTCAATTTGTGTCCGTATGTGCA        60
S3      ------------------------------------------------------------      0
S2      ------------------------------------------------------------      0
S1      ------------------------------------------------------------      0
T1      ------------------------------------------------------------      0

T5      ACTTCATCGTTATATATAGAGGGGGCAATAGCTTCAAGCTAAGCAATTAATATTCAGCAA        71
T4      ------------------------------------------------------------      0
S4      ------------------------------------------------------------      0
S5      ACTACATCGTTATATATAGAGGAGGCAATAGCTTCAAGCTAAGCAATTAATATTCAGAAA        120
S3      ------------------------------------------------------------      0
S2      ------------------------------------------------------------      0
S1      ------------------------------------------------------------      0
T1      ------------------------------------------------------------      0

T5      TGGCTTCCTCAGTTATGTCCTCAGCTGCCGCTGTTGCGACCGGCGCCAATGCTGCTCAAG        131
T4      ------------------------------------------------------------      0
S4      ------------------------------------------------------------      0
S5      TGGCTTCCTCAGTTATGTCCTCAGCTGCCGCTGTTGCCACCGGCGCCAATGCTGCTCAAG        180
S3      ------------------------------------------------------------      0
S2      ------------------------------------------------------------      0
S1      ------------------------------------------------------------      0
T1      ------------------------------------------------------------      0

T5      CCAGTATGGTTGCACCCTTCACTGGTCTCAAGTCCGCAACCTCCTTCCCTGTTTCCAGGA        191
T4      ------------------------------------------------------------      0
S4      ------------------------------------------------------------      0
S5      CCAGTATGGTTGCACCTTTCACTGGTCTCAAGTCCGCAACCTCCTTCCCGGTTTCCAGGA        240
S3      ------------------------------------------------------------      0
S2      ------------------------------------------------------------      0
S1      ------------------------------------------------------------      0
T1      ------------------------------------------------------------      0

T5      AACAAAACCTTGACATTACTTCCATTGCTAGCAACGGCGGAAGAGTTCAATGCATGCAGG        251
T4      ------------------------------------------------------------      0
S4      ------------------------------------------------------------      0
S5      AACAAAACCTTGACATTACTTCCATTGCTAGCAACGGCGGAAGAGTTCAATGCATGCAGG        300
S3      ------------------------------------------------------------      0
S2      ------------------------------------------------------------      0
S1      ------------------------------------------------------------      0
T1      ------------------------------------------------------------      0

T5      TTTGTACTAGAAGTATATAATATCTGGATTGAACTTAGTGTATCTTAACCTGTTTTAGCT        311
T4      ----------------------------TCAA----ATTTATGTGCCCGTCAATTTTTCA        28
S4      ------------------------------------------------------------      0
S5      TTTGTACTAGGAGTATATAATATTTGGATTGAACTTATATACGAGAGTGACATTCTTACT        360
S3      ------------------------------------------------------------      0
S2      ------------------------------------------------------------      0
S1      ------------------------------------------------------------      0
T1      ------------------------------------------------------------      0

T5      TTAGCAGGT----------------------ACCCTATCTTATTTTTTGTCGAT-GATTAC        349
T4      ATCCAACGGCTACGATTCCTCTAAGATGAGGTCATTGCTTGCTTGTGTCCGTTAGATGAG        88
S4      ------CGGTTACGATTCCGCTAAGATGAGGTTATTGCTTGCGTGTGCTCGTTAGATGGG        54
S5      ATATCTAAACCTCTTTTAGCTTTAGGAGATAACTTATCTTATTTTTTTTTGGATGATTAC        420
S3      ----------------------------------GCTTGATTTGTGTCCGTTGATGAG        24
S2      ------------------------------------------TGG-----GGAAGCTAT        12
S1      ------------------------------------------------------------      0
T1      ------------------------------------------------------------      0
```


T5 TTGTAATTAAGGCACTATACTATGTGCCTTAATTTCACCAAATATAT---------- 396
T4 AAAAAGACGTGAA-----ACCTTATCACTATATAT---------- 118
S4 AAAAGGATGTGAA-----ACCTTATCACTATATAT---------- 84
S5 TTGT-A-ATTAAG-----GAACTATACTTGTAATTAAGGCACTATACTATGTTTCGAAAA 473
S3 AAGT-TATATGAA-----GCCTTATCATTATAT------ATATAAGGGGAGCTACAG--- 69
S2 GTGA-AACCTTAT-----CAATTATCATTATAT------ATATCATGGGAGGTACATATC 60
S1 ----------TAGGG---TGGTGGGC 13
T1 ----------TAGGG---TGGTGGGC 13
T5 -------TACTCCCTCCATTTCAATTTAAATTACACAATTTTCCT-----------ATTA 438
T4 --AGCACTCATCACACCCTTGAAAGCAAAGGTCAAGGGAAGCAATAGCTTTAAGCTAAAC 176
S4 --AGCATT---CACACCCTTGAAAGCAAAGGTCAAGGGAAGCAATAGCTTTAAGCTAAAC 139
S5 GAATGACACATTTCTAAATTTGAAAACAATTT-AACTTTAAACTT------TTTATTTTA 526
S3 ----------AGCACTCACCCT------CCTGAAAGC 90
S2 AAATAACCCTCTT-------GAAAGCAAAGGTCAAGGGTAGCAAT------AGCTTTAAG 107
S1 AACTATGCAATGACCATATTGGAAGTTAAAGGAAAAGAGAGAAAGAG----AAATCTTTC 69
T1 AACTATGCAATGACCATCTTGGAAGTTAAGGAAAAGGGAGAAAGAGA----AATCTTTCT 69
T5 GTCCG---TTAAAAAAAGAATGACACGTTTCCAAATTTAAAAACAA---------- 481
T4 AATTACTTTCAACA-----ATGGCTTCGTCTGTGATTTCCTCAGCCGCTGCCGTTGCCAC 231
S4 AATTACTTTCAACAATATAATGGCTTCCTCTGTGATTTCCTCAGCTGTTGCCGTTGCCAC 199
S5 TTCACTTTACTTTTAATGAAAAGCTTTTATAATCATACAAATGGTAT------AACCCCA 580
S3 TAAGCAGTTATTTTCAGAAATGGCTTTCTTAATTATGTCCTCAGCAGCTGCTGTTGCGAC 150
S2 CTTAGAAATTATTTCAGAAATGGCTTCCTCAGTTATGTCCTCAGCAGCTGCTGTTGCGAC 167
S1 TGTCTAAGTGTAATTAACAATGGCTTCCTCAGTTCTTTCCTC---TGCAGCAGTTGCCAC 126
T1 GTCTAAAGTGTAATTAGCAATGGCTTCCTCTGTTCTTTCCTC---TGCAGCAGTTGCCAC 126
T5 ----------TTTAACTTTAAACTTTTCATTTTATCCGCTTTAAGCTT 519
T4 CGGCGCTAATGCGGCTCAAGCCAGTATGGTTGCACCTTTCACT---------- 274
S4 TGGCGCTAATGCTGCTCAAGCCAGTATGGTTGCACCTTTCACT---------- 242
S5 CAAAACTTTTGCCCCTCAAGCTTTTAGAACCACATATTTCAAATACTTTATTTTGGCTTA 640
S3 CGGCACCAATGCTGCTCAAGCCAGCATGATTGCAC----------CCT 188
S2 CGGCGCTAATGCTGCTCAAGCCAGCATGGTTGCAC----------CCT 205
S1 CCGCAGCAATGTTGCTCAAGCTAACATGGTTGCACCTTTC---------- 166
T1 TCGCACCAATGTTGCTCAAGCTAACATGGTTGCACCTTTC---------- 166
T5 T----------TATAATCATACAAATGGCATGGTCCCACAAACCTTTTGCCCCTTAAGC 568
T4 -----GGCCTCAAATCCGCCTACTCCTTCCCTGTTTCC---------- 307
S4 -----GGCCTCAAGTCCGCCTCCTCCTTCCCTGTTACC---------- 275
S5 AACTTTGTGTCAAGTCAAATTACATCATCTAAATTAAAACAGACGGA-----GTATATGC 695
S3 TCACTGGTCTCAAGTCCGCAACCTCCTTCCCTGTTTCC---------- 226
S2 TCACTGGCCTCAAGTCCGCCTCCTCCTTCCCTGTTACC---------- 243
S1 --ACTGGCCTTAAGTCAGCTGCCTCATTCCC---------- 195
T1 --ACTGGTCTTAAGTCAGCTGCCTCATTCCC---------- 195
T5 TTTTAGGATCACAAA--TTTTTTAAATATTTTTTTTTCTTAAACTTTGTGTCAAGT--- 623
T4 -------AGAAAACAAAACCTTGACATTACTTCCATTGCTAGCAATGGTGGAAGAGTTTA 360
S4 -------AGAAAACAAAACCTTGACATTACTTCCATTGCTAGCAATGGTGGAAGAGTCCA 328
S5 ATAGAAGATCACAAAATTCATTGACACCCCAATGTTAGATAGAAAATG---TTATGTTCC 752
S3 -------AGGAAACAAAACCTTGACATTACTTCCATTGCTAGCAATGGTGGAAGAGTTCA 279
S2 -------AGGAAACAAAACCTTGACATTACCTCCATTGCTAGCAATGGTGGAAGAGTTCA 296
S1 TGTTTCAAGGAAGCAAAACCTTGACATCACTTCCATTGCCAGCAACGGCGGAAGAGTGCA 255
T1 TGTTTCAAGGAAGCAAAACCTTGACATCACTTCCATTGCTAGCAATGGTGGAAGAGTGCA 255
T5 -------------C------------AAGCTAAATTATCTAAATTGAAACGGATTGAGTA 658
T4 ATGCATGCAGGTTTGTAGCATATATTATTGTAGTTAGCTTAT--ATAAACTGATAGAGTA 418
S4 ATGCATGCAGGTTTGTATCATATATTATTTT-------TTAT--ATAAATTGATAGTGTA 379
S5 ATAC--TTGTACTGAAGTTAGTTTAAATTTTTTATTNACTTGCAATATACTGATA----- 805
S3 ATGCATGCAGGTTTAATTTGT-----ATATGGATTTAACTTGCAATATACTGATA----- 329
S2 ATGCATGCAGGTTTGTAGTGCATATTATATGGATTTAACTTGCAATATACTGATA----- 351
S1 ATGCATGCAGGTAA---------- 269
T1 ATGCATGCAGGTAA---------- 269
T5 TATGCATAGAAGATCACA-----AAATTCATTGACA------CCCCAATATTAGATTGAA 707
T4 AAGAAATTTTACGTTATATATTGATATATTTTAACCTGGTAATTTGATTTATTTTTCATA 478

```
S4      AAGAAATTTTACGCTATATATTGATATATTTTAACCTGTTAATTTGATTTATTTTTCATG        439
S5      ------------GTACTATAA--AGAACTTTTGCAC------TATCAGTATATGTCACTA        845
S3      ------------GTACTATAA--AGAACTTTTACAC------TATCAGTATATGTTACTA        369
S2      ------------GTACTATAA--AGAACTT-TGCAC------CATCAGTATATGTCACTA        390
S1      ----------------TTTA----TATACAATGACA------GTGCAAAAAATTTTGATA        303
T1      ----------------CTTA----TATACATTCGAC------AATTTTCTTTTTAC----        299
                              *     *

T5      AATGT------TATATTCCATACTTGTACCGTAAATTTTTTATGTGGAATATATATGTAG        761
T4      TTATTAATCCCACTTTTTTATTGTACTTATGAAGTTTATTTTAATTCTTTATATATATAG        538
S4      TTACTAACAATCCCACT--------------------TTTTCTTTAAATTCTTTATATAG        479
S5      CGTTTA---------------TGGAGCTTG-TTTAAATTTTTTTATGTGGGATATATGTAG        889
S3      CGTGGAGT-----AACGTTTATGGAGCTTGTTTA-AATTATTTATGTGGGATATATGTAG        423
S2      AGTGGAGT-----AACATTTATGGAGCTTGTTTAAAATTATTTATTTGGGATACATGTAG        445
S1      CAATTAAT-----GCATCTTAACATGTCATAGCTAAAAATTCTATTTTGGTGGAATATAG        358
T1      -------------------AATTATTGTCATAATTAAAAGTTGTTTTTGGTGGAGTATAG        340
                                                *   *          * ***

T5      GTGTGGCCACCAATTAACAAGAAGAAGTACGAGACTCTCTCATACCTTCCTGATTTGAGC        821
T4      GTGTGGCCACCAATTAACAAGAAGAAGTACGAGACACTCTCATACCTTCCTGATTTGAGC        598
S4      GTGTGGCCACCAATTAACAAGAAGAAGTACGAGACACTCTCATACCTTCCTGATTTGAGC        539
S5      GTGTGGCCACCAATTAACAAGAAGAAGTACGAGACACTCTCATACCTTCCTGATTTGAGC        949
S3      GTGTGGCCACCAATTAACAAGAAGAAGTACGAGACACTCTCATACCTTCCTGATTTGAGC        483
S2      GTGTGGCCACCAATTAACAAGAAGAAGTATGAGACACTCTCATACCTTCCTGATTTGAGC        505
S1      GTGTGGCCACCAATTAACAAGAAGAAGTACGAGACTCTCTCATACCTTCCTGATTTGAGC        418
T1      GTATGGCCCCCATATGGCAAGAAGAAGTACGAAACTCTCTCATACCTTCCCGATTTAAGC        400
        ** ***** ***  *  ************ ** ** ************** ***** ***

T5      GTGGAGCAATTGCTTAGGGAAGTTGAGTACCTTTTGAAAAATGGATGGGTTCCTTGCTTG        881
T4      GAGGAGCAATTGCTTAGGGAAGTTGAATACCTTTTGAAAAATGGATGGGTTCCTTGCTTG        658
S4      GAGGAGCAATTGCTTAGGGAAGTTGAGTACCTTTTGAAAAATGGATGGGTTCCTTGCTTG        599
S5      GAGGAGCAATTGCTTAGGGAAGTTGACTACCTTTTGAAAAATGGATGGGTTCCTTGCTTG        1009
S3      GAGGAGCAATTGCTTAGGGAAGTTGAGTACCTTTTGAAAAATGGATGGGTTCCTTGCTTG        543
S2      CAGGAGCAATTGCTTAGGGAAGTTGATTACCTTTTGAAAAATGGATGGGTTCCTTGCTTG        565
S1      CAGGAGCAATTGCTTAGTGAAGTTGAGTACCTTTTGAAAAATGGATGGGTTCCTTGCTTG        478
T1      GAGGAGCAATTGCTTAGTGAAATTGAGTACCTTTTGAAAAATGGATGGGTTCCTTGTTTG        460
          *************** *** **** ****************************** ***

T5      GAATTCGAGACTGAGGTCAA-CATCTATTCTAAATCTTGCTACTATTATCAAGCATAACT        940
T4      GAATTCGAGACTGAGGTCAAACATCTATTCTAAATCATGCTACTATTATCAAGCATAACT        718
S4      GAATTCGAGACTGAGGTCA-ACATCTATTCTAAATCTTGCTACTATAATCAAGCATATCT        658
S5      GAATTCGAGACTGAGGTCAC-CGTCTATTATAAATCTTGCTACTATTATCAAGCATACCT        1068
S3      GAATTCGAGACTGAGGTCAA-CATCTATTCTAAATCTTGCTACTATTATCAAGCATAACT        602
S2      GAATTCGAGACTGAGGTCAA-CGTCTATTATAAATCTTGCTACTATTATCAAGCATATAT        624
S1      GAATTCGAGACTGAGGTCAATATCTGTTCTAAATTTTGCATACT-CCTTCAATTTTATGC        537
T1      GAATTCGAGACTGAGGTCAATAATTTTGCAT----------ACT-CCCTCTGTTTTATGT        509
        ******************                        ***    **     **

T5      AAC----ATGAATTA-----------------------------------------------CT        953
T4      AACATGAATA----A-----------------------------------------------CT        731
S4      AACATGAATT----A-----------------------------------------------CT        671
S5      AAC----ATGAATTA-----------------------------------------------CT        1081
S3      AACATGATT-----A-----------------------------------------------AT        614
S2      ACCTAACATGAATTA-----------------------------------------------CT        641
S1      GACATTTTTTTCCTTCTATTTGTTCCAAAAAAAAAAGAAGAAGAAGACGACATATTTATA        597
T1      GACTTTTTCTTTTTTATATTTGTTGTCCGAGAAAACAG--------------------ACA        550
          *

T5      CAATCCTAACTATTTTGGGATTATACAT----ATATAGTTGATTAAGT------------        997
T4      CAATCCTAACTAGTTTGGGATTAGACAT----ATATAGTTGATTAAGT------------        775
S4      CAATCCTAACTAGTTTGGGATTATACAG----ATATATTTGATTAAGT------------        715
S5      CAATCCTAACTAATTTGGGATCAAATGT----------------AGTT------------        1113
S3      CAATCCTAACTAATTTGTGA-TCAAACG----------------AGTT------------        645
S2      CAATCCTAACTAATTTGGGATCAAATGT----------------AGTT------------        673
S1      TATTTAGAAAAAATTTAACTTTTAACTTTAATATGTTATTTTGCATGTGCAGCACGGATT        657
T1      TATTTAGAAAAAATTTAACTTTA---------AAATTGCTTAATATGTGTAGCGCGGATT        601
         * *   **  * ***                           *  *

T5      ------------------------------------GAAAGAGGAGTATTA-TCTCATG        1019
T4      ------------------------------------GAAAGAGGAGTATTA-TCTCATG        797
S4      ------------------------------------GAAAGAGGAGTATTA-TCTTATG        737
S5      ------------------------------------GATTAATGGTATTTA-TCTCATG        1135
S3      ------------------------------------GATTAATGGTATTTA-TATCGTG        667
S2      ------------------------------------GATTAATGGTATTTA-TTTCATG        695
```

```
S1      TGTCTACCGTGAAAACAACAAGTCACCAGGATACTATGATGGCAGGTCAGTACCAAATCC        717
T1      TGTCTACCGTGAAAACAACAAGTCACCAGGATACTATGACGGCAGGTGAGTCACAATTAT        661
                                          **     *     *

T5      TTAATGTTTTGTTTATCTTGTGGATATGTGCAGC----------------------------      1053
T4      TTAATGTTTTGTTTATCTTGTGGATATGCGCAGC----------------------------      831
S4      TTAATGTTTTGTTTATCTTGTGGATATGTGCAGC----------------------------      771
S5      TTAATGTTTTGTTTATCTTGTGGATATGTACAGC----------------------------      1169
S3      TTAATGTTTTGTTTATCTTGTGGATATGTGCAGC----------------------------      701
S2      TTAATGTTTTGTTTATCTTGTGGATATGTACAGC----------------------------      729
S1      ACAATATTTTAGTTAAATCATAGATATATGTACGCAGATCTTAAAAAGAAGTACAAAAAT        777
T1      TTAGT--------TAAATCATGAATAT-----------------------------------      680
          * *        **  *  *  ****

T5      --------------AC----GGATTCGTCTACCGTGAGAACAACAAGTCACCAGGTTACT        1095
T4      --------------AC----GGATTCGTCTACCGTGAGAATAACAAGTCACCAGGTTACT        873
S4      --------------AC----GGATTCGTCTACCGTGAGAACAACAAGTCACCAGGGTACT        813
S5      --------------AC----GGATTCGTCTACCGTGAGAACAACAAGTCACCAGGATATT        1211
S3      --------------AC----GGATTCGTCTACCGCGAGAACAATAAATCACCGGGGTACT        743
S2      --------------AC----GGATTCGTCTACCGTGAGAACAATAAGTCACCAGGGTACT        771
S1      CATGTTTCGAGACCATGTTTTAGAAGAACGAAATCTTGAATTTTAAATCTTGAATCTGGT        837
T1      ------------------------------AAATGTTAAAT---------TTTAAATCTT        701
                                      *       **               *  *

T5      ACGATGGAAGGTACTGGACCATGTGGAAGTTGCCCATGTTCGGGTGCACTGATGCCACTC        1155
T4      ACGATGGAAGGTACTGGACCATGAGGAAGTTGCCCATGTTCGGGTGCACTGATGCCACTC        933
S4      ACGATGGTAGATACTGGACCATGTGGAAGTTGCCCATGTTCGGGTGCACTGATGCCACTC        873
S5      ATGATGGTAGATACTGGACCATGTGGAAGTTGCCCATGTTCGGGTGCACTGATGCCACTC        1271
S3      ATGATGGCAGATACTGGACCATGTGGAAGTTGCCCATGTTCGGGTGCACTGATGCCACTC        803
S2      ACGATGGTAGATACTGGACCATGTGGAAGTTGCCCATGTTCGGGTGCACTGATGCCACTC        831
S1      ACTGATGCAGATACTGGACCATGTGGAAGCTACCTATGTTCGGATGCACTGATGCCACCC        897
T1      GAATGCGCAGATACTGGACCATGTGGAAGCTGCCCATGTTTGGGTGCACTGATGCCACCC        761
              * ** ************ ***** * ** ***** ** ************** *

T5      AGGTCTTGGCTGAGGTCGAGGAGGCTAAGAAGGCTTACCCACAAGCCTGGATCAGAATCA        1215
T4      AGGTCTTGGCTGAGGTCGAGGAGGCAAAGAAGGCTTACCCACAAGCCTGGATCAGAATCA        993
S4      AGGTCTTGGCTGAGGTAGAGGAGGCAAAGAAGGCTTACCCACAAGCCTGGATCAGAATCA        933
S5      AGGTCTTGGCTGAGGTAGAGGAGGCAAAGAAGGCTTACCCACAAGCCTGGATCAGAATCA        1331
S3      AGGTATTGGCTGAGGTAGAGGAGGCAAAGAAGGCTTACCCACAAGCCTGGATCAGAATAA        863
S2      AGGTCTTGGCTGAGGTAGAGGAGGCAAAGAAGGCTTACCCACAGGCCTGGATCAGAATCA        891
S1      AAGTGTTGGCTGAGGTGGAAGAGGCGAAGAAGGCATACCCACAGGCCTGGATCCGTATCA        957
T1      AAGTGTTAGCTGAGGTGGGAGAGGCGAAGAAGGCATACCCAGAGGCCTGGATCCGTATCA        821
        * ** ** ******** *  ***** ******** ****** * ********* * ** *

T5      TTGGATTCGACAACGTCCGTCAAGTGCAGTGCATCAGTTTTATTGCCTACAAGCCTGAAG        1275
T4      TTGGATTCGACAACGTCCGTCAAGTGCAATGCATCAGTTTCATCGCCTACAAGCCCGCAG        1053
S4      TTGGATTCGACAACGTCCGTCAAGTGCAATGCATCAGTTTCATCGCCTACAAGCCCGAAG        993
S5      TTGGATTTGATAACGTCCGTCAAGTGCAGTGCATCAGTTTCATCGCCTACAAGCCCGAAG        1391
S3      TTGGATTCGACAATGTCCGTCAAGTGCAGTGCATTAGTTTCATCGCCTACAAGCCCGAAG        923
S2      TTGGATTCGACAACGTCCGTCAAGTGCAATGCATCAGTTTCATCGCCTACAAGCCCGAAG        951
S1      TTGGATTCGACAACGTGCGTCAAGTGCAGTGCATCAGTTTCATTGCCTACAAGCCAGAAG        1017
T1      TTGGATTCGACAACGTGCGTCAAGTGCAGTGCATCAGTTTCATTGCCTACAAGCCTGAAG        881
        ******* ** ** ** *********** ***** ***** ** *********** * **

T5      GCTACTAAAATCTCCATTTTAAAGACAGCTTAC--CTATGTCTTCGGGGGAAGT-----        1328
T4      GCTACTAAAATCTCCATTTTTAAGACAACTTACC--GTATGTATTCAGGGGAAGTTTGTT        1111
S4      GCTACTAAAATCTCCATTTTTAAGACAACTTACC--CCATGTCTTCGGGGGAAGTTTGTT        1051
S5      GCTACTAAAATCTCTATTTTTAAGACAACTTACC--GTATGTATTCAGGGGAAGTTTGTT        1449
S3      GCTACTAAAATCTCCATTTTTAAGACAACTTACT--GTATGTATTCAGGGAAAGTTTGTT        981
S2      GCTACTAAAATCTCCGTTTTTAAGACAACTTACC--GTATGGATTCAGGGGAAGCTTGTT        1009
S1      GCTACTAAGTTTCATATTAGGACAACTTACCCTATTGTCTGACTTTAGGGGCAGTTTGTT        1077
T1      GCTACTAAGTTACATATTAGGACAACTTCCCTATTGTCTTGTCTTTAGGGGTTGTGTTGT        941
        ********  *     **   *  **            **  **  ***   *

T5      ------------------------------------------------------------        1328
T4      TGAATTCTCCTTGTGTTTTTCCCCGGAGAAACTGTTTTGGTTTTCC--------------        1157
S4      ------------------------------------------------------------        1051
S5      TGAATTCTCCTTGAGTTTTTCCCCGGAAAAACTGTTTTGGTTTTCTTTTGTTTTAATTCC        1509
S3      TGCA----------------------------------------------------------        985
S2      TGAATCTCCTTGGGC-----------------------------------------------        1024
S1      TGAA----------------------------------------------------------        1081
T1      TTTAATTTTTTTTTACTTCTTCCCACAAAAACTGTTTATGT--------------TTCC        986
```

```
T5      ------------------------------------------------------------        1328
T4      ------------------------------------------------------------        1157
S4      ------------------------------------------------------------        1051
S5      TTCTTTCTATTCGGTATATATTTTTGAATTCCACTCAAGTTTATGAGAACTAATAATAAT        1569
S3      ------------------------------------------------------------        985
S2      ------------------------------------------------------------        1024
S1      ------------------------------------------------------------        1081
T1      TTCTTTCTATTCGGTGTATGTTTTTCGATTCCTACCAA-GTTATGAGACCTAATAATTAT        1045

T5      ------------------------------------------------------------        1328
T4      ------------------------------------------------------------        1157
S4      ------------------------------------------------------------        1051
S5      CATTTGTTTCTTTACTAATTTGAAACATGTTCTCTGGCATAAGTCAACATCCGGTCAACT        1629
S3      ------------------------------------------------------------        985
S2      ------------------------------------------------------------        1024
S1      ------------------------------------------------------------        1081
T1      GATTTGGTGCTTTGTTTGTATAATATTTTTGTTTC-ACATTCTTGTGC---CG-------        1094

T5      ---------------------------------        1328
T4      ---------------------------------        1157
S4      ---------------------------------        1051
S5      TTACCAACTACTCTTTCTGTCCCGGAGTCTTG        1661
S3      ---------------------------------        985
S2      ---------------------------------        1024
S1      ---------------------------------        1081
T1      ---------------------------------        1094
```

Appendix B Figure 1. MSA of eight full-length *rbcS* isoforms. Full-length sequences were verified by sequencing PCR amplicons using gene-specific primers (**bold**) (Table 2–1). Most isoforms are encoded by three exons (red) except *rbcS-T1* and *rbcS-S1*, which are encoded by four exons.

Coding sequences

```
S3      ATGGCTTTCTTAATTATGTCCTCAGCAGCTGCTGTTGCGACCGGCACCAATGCTGCTCAA        60
S5      ATGGCTTCCTCAGTTATGTCCTCAGCTGCCGCTGTTGCCACCGGCGCCAATGCTGCTCAA        60
T5      ATGGCTTCCTCAGTTATGTCCTCAGCTGCCGCTGTTGCGACCGGCGCCAATGCTGCTCAA        60
T4      ATGGCTTCGTCTGTGATTTCCTCAGCCGCTGCCGTTGCCACCGGCGCTAATGCGGCTCA-        59
S2      ATGGCTTCCTCAGTTATGTCCTCAGCAGCTGCTGTTGCGACCGGCGCTAATGCTGCTCAA        60
S4      ATGGCTTCCTCTGTGATTTCCTCAGCTGTTGCCGTTGCCACTGGCGCTAATGCTGCTCAA        60
S1      ATGGCTTCCTCAGTTCTTTCCTCTGCAGC---AGTTGCCACCCGCAGCAATGTTGCTCAA        57
T1      ATGGCTTCCTCTGTTCTTTCCTCTGCAGC---AGTTGCCACTCGCACCAATGTTGCTCAA        57
        *******  *   *  * ***** ** *     ***** **  **   ****  *****

S3      GCCAGCATGATTGCACCCTTCACTGGTCTCAAGTCCGCAACCTCCTTCCCTGTTTCCAGG        120
S5      GCCAGTATGGTTGCACCTTTCACTGGTCTCAAGTCCGCAACCTCCTTCCCGGTTTCCAGG        120
T5      GCCAGTATGGTTGCACCCTTCACTGGTCTCAAGTCCGCAACCTCCTTCCCTGTTTCCAGG        120
T4      GCCAGTATGGTTGCACCTTTCACTGGCCTCAAATCCGCCTACTCCTTCCCTGTTTCCAGA        119
S2      GCCAGCATGGTTGCACCCTTCACTGGCCTCAAGTCCGCCTCCTCCTTCCCTGTTACCAGG        120
S4      GCCAGTATGGTTGCACCTTTCACTGGCCTCAAGTCCGCCTCCTCCTTCCCTGTTACCAGA        120
S1      GCTAACATGGTTGCACCTTTCACTGGCCTTAAGTCAGCTGCCTCATTCCCTGTTTCAAGG        117
T1      GCTAACATGGTTGCACCTTTCACTGGTCTTAAGTCAGCTGCCTCATTCCCTGTTTCAAGG        117
        ** *  *** ******* ******** ** ** ** **   *** ***** *** * **

S3      AAACAAAACCTTGACATTACTTCCATTGCTAGCAATGGTGGAAGAGTTCAATGCATGCAG        180
S5      AAACAAAACCTTGACATTACTTCCATTGCTAGCAACGGCGGAAGAGTTCAATGCATGCAG        180
T5      AAACAAAACCTTGACATTACTTCCATTGCTAGCAACGGCGGAAGAGTTCAATGCATGCAG        180
T4      AAACAAAACCTTGACATTACTTCCATTGCTAGCAATGGTGGAAGAGTTTAATGCATGCAG        179
S2      AAACAAAACCTTGACATTACCTCCATTGCTAGCAATGGTGGAAGAGTTCAATGCATGCAG        180
S4      AAACAAAACCTTGACATTACTTCCATTGCTAGCAATGGTGGAAGAGTCCAATGCATGCAG        180
S1      AAGCAAAACCTTGACATCACTTCCATTGCCAGCAACGGCGGAAGAGTGCAATGCATGCAG        177
T1      AAGCAAAACCTTGACATCACTTCCATTGCTAGCAATGGTGGAAGAGTGCAATGCATGCAG        177
        ** ************** ** ******** ***** ** ********  ***********

S3      GTGTGGCCACCAATTAACAAGAAGAAGTACGAGACACTCTCATACCTTCCTGATTTGAGC        240
S5      GTGTGGCCACCAATTAACAAGAAGAAGTACGAGACACTCTCATACCTTCCTGATTTGAGC        240
T5      GTGTGGCCACCAATTAACAAGAAGAAGTACGAGACTCTCTCATACCTTCCTGATTTGAGC        240
```

```
T4      GTGTGGCCACCAATTAACAAGAAGAAGTACGAGACACTCTCATACCTTCCTGATTTGAGC        239
S2      GTGTGGCCACCAATTAACAAGAAGAAGTATGAGACACTCTCATACCTTCCTGATTTGAGC        240
S4      GTGTGGCCACCAATTAACAAGAAGAAGTACGAGACACTCTCATACCTTCCTGATTTGAGC        240
S1      GTGTGGCCACCAATTAACAAGAAGAAGTACGAGACTCTCTCATACCTTCCTGATTTGAGC        237
T1      GTATGGCCCCCATATGGCAAGAAGAAGTACGAAACTCTCTCATACCTTCCCGATTTAAGC        237
        ** ***** ***  *  ************ ** ** ************** ***** ***

S3      GAGGAGCAATTGCTTAGGGAAGTTGAGTACCTTTTGAAAAATGGATGGGTTCCTTGCTTG        300
S5      GAGGAGCAATTGCTTAGGGAAGTTGACTACCTTTTGAAAAATGGATGGGTTCCTTGCTTG        300
T5      GTGGAGCAATTGCTTAGGGAAGTTGAGTACCTTTTGAAAAATGGATGGGTTCCTTGCTTG        300
T4      GAGGAGCAATTGCTTAGGGAAGTTGAATACCTTTTGAAAAATGGATGGGTTCCTTGCTTG        299
S2      CAGGAGCAATTGCTTAGGGAAGTTGATTACCTTTTGAAAAATGGATGGGTTCCTTGCTTG        300
S4      GAGGAGCAATTGCTTAGGGAAGTTGAGTACCTTTTGAAAAATGGATGGGTTCCTTGCTTG        300
S1      CAGGAGCAATTGCTTAGTGAAGTTGAGTACCTTTTGAAAAATGGATGGGTTCCTTGCTTG        297
T1      GAGGAGCAATTGCTTAGTGAAATTGAGTACCTTTTGAAAAATGGATGGGTTCCTTGTTTG        297
          *************** *** **** ****************************** ***

S3      GAATTCGAGACTGAGCACGGATTCGTCTACCGCGAGAACAATAAATCACCGGGGTACTAT        360
S5      GAATTCGAGACTGAGCACGGATTCGTCTACCGTGAGAACAACAAGTCACCAGGATATTAT        360
T5      GAATTCGAGACTGAGCACGGATTCGTCTACCGTGAGAACAACAAGTCACCAGGTTACTAC        360
T4      GAATTCGAGACTGAGCACGGATTCGTCTACCGTGAGAATAACAAGTCACCAGGTTACTAC        359
S2      GAATTCGAGACTGAGCACGGATTCGTCTACCGTGAGAACAATAAGTCACCAGGGTACTAC        360
S4      GAATTCGAGACTGAGCACGGATTCGTCTACCGTGAGAACAACAAGTCACCAGGGTACTAC        360
S1      GAATTCGAGACTGAGCACGGATTTGTCTACCGTGAAAACAACAAGTCACCAGGATACTAT        357
T1      GAATTCGAGACTGAGCGCGGATTTGTCTACCGTGAAAACAACAAGTCACCAGGATACTAT        357
        **************** ****** ******** ** ** ** ** ***** ** ** **

S3      GATGGCAGATACTGGACCATGTGGAAGTTGCCCATGTTCGGGTGCACTGATGCCACTCAG        420
S5      GATGGTAGATACTGGACCATGTGGAAGTTGCCCATGTTCGGGTGCACTGATGCCACTCAG        420
T5      GATGGAAGGTACTGGACCATGTGGAAGTTGCCCATGTTCGGGTGCACTGATGCCACTCAG        420
T4      GATGGAAGGTACTGGACCATGAGGAAGTTGCCCATGTTCGGGTGCACTGATGCCACTCAG        419
S2      GATGGTAGATACTGGACCATGTGGAAGTTGCCCATGTTCGGGTGCACTGATGCCACTCAG        420
S4      GATGGTAGATACTGGACCATGTGGAAGTTGCCCATGTTCGGGTGCACTGATGCCACTCAG        420
S1      GATGGCAGATACTGGACCATGTGGAAGCTACCTATGTTCGGATGCACTGATGCCACCCAA        417
T1      GACGGCAGATACTGGACCATGTGGAAGCTGCCCATGTTTGGGTGCACTGATGCCACCCAA        417
        ** ** ** ************ ***** * ** ***** ** ************** **

S3      GTATTGGCTGAGGTAGAGGAGGCAAAGAAGGCTTACCCACAAGCCTGGATCAGAATAATT        480
S5      GTCTTGGCTGAGGTAGAGGAGGCAAAGAAGGCTTACCCACAAGCCTGGATCAGAATCATT        480
T5      GTCTTGGCTGAGGTCGAGGAGGCTAAGAAGGCTTACCCACAAGCCTGGATCAGAATCATT        480
T4      GTCTTGGCTGAGGTCGAGGAGGCAAAGAAGGCTTACCCACAAGCCTGGATCAGAATCATT        479
S2      GTCTTGGCTGAGGTAGAGGAGGCAAAGAAGGCTTACCCACAGGCCTGGATCAGAATCATT        480
S4      GTCTTGGCTGAGGTAGAGGAGGCAAAGAAGGCTTACCCACAAGCCTGGATCAGAATCATT        480
S1      GTGTTGGCTGAGGTGGAAGAGGCGAAGAAGGCATACCCACAGGCCTGGATCCGTATCATT        477
T1      GTGTTAGCTGAGGTGGGAGAGGCGAAGAAGGCATACCCAGAGGCCTGGATCCGTATCATT        477
        ** ** ******** *  ***** ******** ****** * ********* * ** ***

S3      GGATTCGACAATGTCCGTCAAGTGCAGTGCATTAGTTTCATCGCCTACAAGCCCGAAGGC        540
S5      GGATTTGATAACGTCCGTCAAGTGCAGTGCATCAGTTTCATCGCCTACAAGCCCGAAGGC        540
T5      GGATTCGACAACGTCCGTCAAGTGCAGTGCATCAGTTTTATTGCCTACAAGCCTGAAGGC        540
T4      GGATTCGACAACGTCCGTCAAGTGCAATGCATCAGTTTCATCGCCTACAAGCCCGCAGGC        539
S2      GGATTCGACAACGTCCGTCAAGTGCAATGCATCAGTTTCATCGCCTACAAGCCCGAAGGC        540
S4      GGATTCGACAACGTCCGTCAAGTGCAATGCATCAGTTTCATCGCCTACAAGCCCGAAGGC        540
S1      GGATTCGACAACGTGCGTCAAGTGCAGTGCATCAGTTTCATTGCCTACAAGCCAGAAGGC        537
T1      GGATTCGACAACGTGCGTCAAGTGCAGTGCATCAGTTTCATTGCCTACAAGCCTGAAGGC        537
        ***** ** ** ** *********** ***** ***** ** *********** * ****

S3      TAATAA      546
S5      TACTAA      546
T5      TACTAA      546
T4      TACTAA      545
S2      TACTAA      546
S4      TACTAA      546
S1      TACTAA      543
T1      TACTAA      543
        ** ***
```

Appendix B Figure 2. MSA of the coding sequences of eight full-length *rbcS* isoforms. Coding sequences were verified by Sanger sequencing of RT-PCR products with gene-specific primers (Table 2–1). MSA was performed using Clustal Omega (EMBL-EBI) and was used to determine the percent similarity between each isoform (Fig. 3–2).

SSU-4 670-bp deletion (*rbcS-T1*)

```
rbcS_T1_WT          1 T---AGGGTGG-TGGGCAACTATGCAATGACCATCTTGGAAGTTAAGGAA     46
                      |   ||||||| |||||||||||||||||||||||||||||||||||||
reversecomp         1 TTTTAGGGTGGGTGGGCAACTATGCAATGACCATCTTGGAAGTTAAGGAA     50

rbcS_T1_WT         47 AAGGGAGAAAGAGAAATCTTTCTGTCTAAAGTGTAATTAGCAATGGCTTC     96
                      ||||||||||||||||||||||||||||||||||||||||||||||||||||
reversecomp        51 AAGGGAGAAAGAGAAATCTTTCTGTCTAAAGTGTAATTAGCAATGGCTTC    100

rbcS_T1_WT         97 CTCTGTTCTTTCCTCTGCAGCAGTTGCCACTCGCACCAATGTTGCTCAAG    146
                      ||||||||||||||||||||||||||||||||||||||||||||||||||||
reversecomp       101 CTCTGTTCTTTCCTCTGCAGCAGTTGCCACTCGCACCAATGTTGCTCAAG    150

rbcS_T1_WT        147 CTAACATGGTTGCACCTTTCACTGGTCTTAAGTCAGCTGCCTCATTCCCT    196
                      ||
reversecomp       151 CT----------------------------------------------    152

rbcS_T1_WT        197 GTTTCAAGGAAGCAAAACCTTGACATCACTTCCATTGCTAGCAATGGTGG    246

reversecomp       153 ----------------------------------------------    152

rbcS_T1_WT        247 AAGAGTGCAATGCATGCAGGTAACTTATATACATTCGACAATTTTCTTTT    296

reversecomp       153 ----------------------------------------------    152

rbcS_T1_WT        297 TACAATTATTGTCATAATTAAAAGTTGTTTTTGGTGGAGTATAGGTATGG    346

reversecomp       153 ----------------------------------------------    152

rbcS_T1_WT        347 CCCCCATATGGCAAGAAGAAGTACGAAACTCTCTCATACCTTCCCGATTT    396

reversecomp       153 ----------------------------------------------    152

rbcS_T1_WT        397 AAGCGAGGAGCAATTGCTTAGTGAAATTGAGTACCTTTTGAAAAATGGAT    446

reversecomp       153 ----------------------------------------------    152

rbcS_T1_WT        447 GGGTTCCTTGTTTGGAATTCGAGACTGAGGTCAATAATTTTGCATACTCC    496

reversecomp       153 ----------------------------------------------    152

rbcS_T1_WT        497 CTCTGTTTTATGTGACTTTTTCTTTTTTATATTTGTTGTCCGAGAAAACA    546

reversecomp       153 ----------------------------------------------    152

rbcS_T1_WT        547 GACATATTTAGAAAAAATTTAACTTTAAAATTGCTTAATATGTGTAGCGC    596

reversecomp       153 ----------------------------------------------    152

rbcS_T1_WT        597 GGATTTGTCTACCGTGAAAACAACAAGTCACCAGGATACTATGACGGCAG    646

reversecomp       153 ----------------------------------------------    152

rbcS_T1_WT        647 GTGAGTCACAATTATTTAGTTAAATCATGAATATAAATGTTAAATTTTAA    696

reversecomp       153 ----------------------------------------------    152

rbcS_T1_WT        697 ATCTTGAATGCGCAGATACTGGACCATGTGGAAGCTGCCCATGTTTGGGT    746

reversecomp       153 ----------------------------------------------    152

rbcS_T1_WT        747 GCACTGATGCCACCCAAGTGTTAGCTGAGGTGGGAGAGGCGAAGAAGGCA    796

reversecomp       153 ----------------------------------------------    152

rbcS_T1_WT        797 TACCCAGAGGCCTGGATCCGTATCATTGGATTCGACAACGTGCGTCAAGT    846
                                                 ||||||||||||||||||||||||||||
reversecomp       153 ----------------------TCATTGGATTCGACAACGTGCGTCAAGT    180

rbcS_T1_WT        847 GCAGTGCATCAGTTTCATTGCCTACAAGCCTGAAGGCTACTAAGTTACAT    896
                      ||||||||||||||||||||||||||||||||||||||||||||||||||||
reversecomp       181 GCAGTGCATCAGTTTCATTGCCTACAAGCCTGAAGGCTACTAAGTTACAT    230
```

```
rbcS_T1_WT       897 ATTAGGACAACTT-CCCTATTGTCTTGTCTTTAGGGGTTGTGTTGTTTTA    945
                     |||||||||||||  |||||||||||||||||||||||||||||||||||
reversecomp      231 ATTAGGACAACTTACCCTATTGTCTTGTCTTTAGGGGTTGTGTTGTTTTA    280

rbcS_T1_WT       946 ATTTTTTTTTACTTCTTCCCACAAAAACTGTTTATGTTTCCTTCTTTCTA    995
                     ||||||||||||||||||||||||||||||||||||||||||||||||||||
reversecomp      281 ATTTTTTTTTACTTCTTCCCACAAAAACTGTTTATGTTTCCTTCTTTCTA    330

rbcS_T1_WT       996 TTCGGTGTATGTTTTTCGATTCCTACCAAGTTATGAGACCTAATAATTAT   1045
                     ||||||||||||||||||||||||||||||||||||||||||||||||||||
reversecomp      331 TTCGGTGTATGTTTTTCGATTCCTACCAAGTTATGAGACCTAATAATTAT    380

rbcS_T1_WT      1046 GATTTGGTGCTTTGTTTGTATAATATTTTTGTTTCACATTCTTGTGCCG   1094
                     ||||||||||.|.||| |.||.
reversecomp      381 GATTTGGTGCGTGGTT-GAATT--------------------------    401
```

Appendix B Figure 2. Multiple sequence alignment of *rbcS-T1* from *SSU-4* and wild-type (WT). Sequences obtained from Sanger sequencing of PCR products amplified using gene-specific primers. Multiple sequence alignment shows a 670-bp deletion between the two gRNA sites (underlined) in *rbcS-T1* from *SSU-4* in the T0 generation.